矩阵分析

王朝瑞　史荣昌　编著

北京理工大学出版社

内 容 提 要

全书共十二章，主要介绍矩阵理论中的基本内容。本书对λ-矩阵，函数矩阵，酉矩阵，Hermite矩阵，正规矩阵和矩阵偶，矩阵分解，矩阵函数等内容作了较为详细的讨论，还讨论了矩阵方程和广义逆矩阵，书中每章均附有习题，有助于读者巩固和提高。

本书可以作为高等工科院校研究生教材，也可供高年级学生和有关工程技术人员参考。

矩 阵 分 析

王朝瑞 史荣昌 编著

*

北京理工大学出版社出版

新华书店北京发行所发行 各地新华书店经售

北京理工大学出版社印刷厂印刷

*

787×1092毫米 16开本 18.625印张 460千字

1989年8月第一版 1989年8月第一次印刷

ISBN 7-81013-230-×/O·38

印数：1—3000册 定价：4.25元

前　言

本书是作者在为北京理工大学研究生院讲授“矩阵分析”时所编讲义的基础上修改而成。

由于自然科学和工程技术的迅速发展，特别由于计算机的普遍使用，使得矩阵理论得到日益广泛的应用。目前，国内大多数工科院校均为研究生开设了矩阵论课，对矩阵论课的教学提出了新的要求。我们认为，一本较适用的教材一方面应当具有一定的理论深度和广度，以满足近代自然科学和工程技术的需要；另一方面，又应深入浅出，力求做到叙述简洁易懂。

矩阵理论的内容十分丰富，在一本书中包括矩阵理论的全部内容几乎是不可能的，本书介绍的是矩阵理论的基本内容，但它们又是工程技术中经常用到的。全书共十二章，第一章介绍线性空间和线性变换；第二章和第三章讨论以多项式为元素和以函数为元素的矩阵；第四章讨论酉矩阵和Hermite矩阵；第五章讨论正规矩阵和矩阵偶；第六章讨论矩阵分解；第七章和第八章讨论矩阵范数和矩阵序列与矩阵级数的收敛性；第九章讨论矩阵函数，以矩阵函数的各种表示为主要内容；第十章和第十一章讨论矩阵微分方程和矩阵代数方程；第十二章讨论广义逆矩阵。

由于作者水平有限，不妥之处实属难免，请读者批评指正。

编著者

1987.12 于北京

目　录

第一章　线性空间与线性变换

第二章　λ-矩阵与标准形

第三章　函数矩阵

第四章　酉矩阵　Hermite 矩阵

第五章 正规矩阵与矩阵偶的标准形

第六章 矩阵的分解

第七章 范数 测度

第八章 矩阵序列和矩阵级数

第九章 矩阵函数

第一章　线性空间与线性变换

线性空间与线性变换是线性代数中的两个重要内容。为了后面讨论方便起见，本章介绍线性空间与线性变换的基本概念。主要内容有：线性空间，基变换与坐标变换，子空间与子空间的运算，线性变换，线性变换与矩阵，特征值，特征向量，化矩阵为对角形。

§1.1　线 性 空 间

线性空间是线性代数最基本的概念之一，这节我们研究它的定义、维数、基底与坐标。

定义 1.1.1 设 V 是一个非空集合，F 是一个数域，在集合 V 的元素之间定义了称为加法的运算。也就是说给出了一个法则，对于 V 中任意两个元素 $\boldsymbol{x}$ 与 $\boldsymbol{y}$，在 V 中都有唯一的元素 $\boldsymbol{z}$ 与它们相对应，称为 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的和，记为 $\boldsymbol{z}=\boldsymbol{x}+\boldsymbol{y}$。并且加法运算满足下面四条法则：

（1）　交换律　$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$。

（2）　结合律　$\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})=(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$。

（3）　零元素　在 V 中有一元素 $\mathbf{0}$（称作零元素），对于 V 中任一元素 $\boldsymbol{x}$ 都有

$$\boldsymbol{x}+\mathbf{0}=\boldsymbol{x}$$

（4）　负元素　对于 V 中每一个元素 $\boldsymbol{x}$，都有 V 的元素 $\boldsymbol{y}$，使得

$$\boldsymbol{x}+\boldsymbol{y}=\mathbf{0}$$

在集合 V 的元素与数域 F 中的数之间还定义了一种运算，叫做**数量乘法**。也就是说，对于 V 中任一元素 $\boldsymbol{x}$ 与 F 中任一数 k，在 V 中有唯一的一个元素 $\boldsymbol{u}$ 与它们对应，称为 k 与 $\boldsymbol{x}$ 的数量乘积，记为 $\boldsymbol{u}=k\boldsymbol{x}$，并且数量乘法与加法满足下面四条法则：

（1）　$1\cdot\boldsymbol{x}=\boldsymbol{x}$

（2）　$k(l\boldsymbol{x})=(kl)\boldsymbol{x}$

（3）　$(k+l)\boldsymbol{x}=k\boldsymbol{x}+l\boldsymbol{x}$

（4）　$k(\boldsymbol{x}+\boldsymbol{y})=k\boldsymbol{x}+k\boldsymbol{y}$

其中 k、l 表示数域 F 中的任意数，$\boldsymbol{x}$、$\boldsymbol{y}$ 表示 V 中任意元素。

这样 V 称为数域 F 上的**线性空间**。一般情况下，F 是实数域 R 或复数域 C。

例 1.1.1 元素属于数域 F 的 $m\times n$ 阶矩阵，按矩阵的加法和矩阵与数的数量乘法，构成数域 P 上的一个线性空间，用 $F^{m\times n}$ 表示。$m\times n$ 阶实数矩阵所成的空间用 $R^{m\times n}$ 表示，$m\times n$阶复数矩阵所成的空间用 $C^{m\times n}$ 表示。

例 1.1.2 数域 F 按照本身的加法与乘法，构成一个自身上的线性空间。

例 1.1.3 系数都属于数域 F 的一元多项式的全体，按通常多项式加法和数与多项式的

乘法，构成一个数域 F 上的线性空间。

如果只考虑其中次数小于 n 的多项式全体（包括零多项式），构成一个线性空间，用 $F[x]_n$ 表示。

例 1.1.4 线性齐次方程组

$$A\boldsymbol{x}=0$$

的所有解的集合构成一个线性空间。

例 1.1.5 空间解析几何中所有过原点的向量集合按通常的向量加法和数与向量的乘法，构成一个线性空间。

例 1.1.6 定义在闭区间 $[a,b]$ 上连续函数全体按通常函数的加法和数与函数的乘法，构成一个线性空间。

线性空间 V 的元素 $\boldsymbol{x}$，$\boldsymbol{y}$，$\boldsymbol{z}$，…称为向量。这里所谓向量比几何中的向量涵义要广泛得多。

定义 1.1.2 设 V 是数域 F 上的一个线性空间，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r(r\geqslant 1)$ 是 V 中一组向量，k_1，k_2，…，k_r 是数域 F 中的数，那么向量

$$\boldsymbol{x}=k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_r\boldsymbol{x}_r$$

称为向量 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 的一个线性组合，有时我们也说向量 $\boldsymbol{x}$ 可以用向量组 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 线性表出。

定义 1.1.3 线性空间 V 中向量 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ $(r\geqslant 1)$ 称为线性相关，如果在数域 F 中有 r 个不全为零的数 k_1，k_2，…，k_r 使得

$$k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_r\boldsymbol{x}_r=0 \qquad (1\text{-}1)$$

如果向量 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 不线性相关，就称为线性无关。换句话说，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 称为线性无关，如果等式（1-1）只有在 $k_1=k_2=\cdots=k_r=0$ 时才成立。

定理 1.1.1 设向量组 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 线性无关，但 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$，$\boldsymbol{y}$ 线性相关，则 $\boldsymbol{y}$ 可以用 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 线性表出，且表出是唯一的。

〔证明〕 由定理条件知，存在不全为零的数 k_1，k_2，…，k_r，l 使得

$$k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_r\boldsymbol{x}_r+l\boldsymbol{y}=0 \qquad (1\text{-}2)$$

于是 $l\neq 0$，否则的话，若 $l=0$，则有

$$k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_r\boldsymbol{x}_r=0$$

因为 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 线性无关，所以

$$k_1=k_2=\cdots=k_r=0$$

这与 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$，$\boldsymbol{y}$ 线性相关矛盾！于是 $l\neq 0$，故得

$$\begin{aligned}\boldsymbol{y}&=-\frac{k_1}{l}\boldsymbol{x}_1-\frac{k_2}{l}\boldsymbol{x}_2-\cdots-\frac{k_r}{l}\boldsymbol{x}_r\\&=c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2+\cdots+c_r\boldsymbol{x}_r\end{aligned}$$

此即 $\boldsymbol{y}$ 可以用 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$ 线性表出。现证表出是唯一的。

设 $\boldsymbol{y}$ 有两个表示式，即

$$\boldsymbol{y}=c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2+\cdots+c_r\boldsymbol{x}_r$$

$$\boldsymbol{y}=c_1'\boldsymbol{x}_1+c_2'\boldsymbol{x}_2+\cdots+c_r'\boldsymbol{x}_r$$

由上面的等式，有

$$(c_1-c_1')\boldsymbol{x}_1+(c_2-c_2')\boldsymbol{x}_2+\cdots+(c_r-c_r')\boldsymbol{x}_r=0$$

因为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$ 线性无关，所以

$$c_i-c_i'=0 \quad (i=1,2,\cdots,r)$$

即
$$c_i=c_i' \quad (i=1,2,\cdots,r) \tag{1-3}$$ ▮

符号"▮"表示证明结束。

定义 1.1.4 如果在线性空间 V 中有 n 个线性无关的向量，但是没有更多数目的线性无关的向量，那么 V 就称为 $\boldsymbol{n}$ **维的**。如果在 V 中可以找到任意多个线性无关的向量，那么 V 就称为**无限维的**。V 的维数记作 $\dim V$。

按照这个定义，不难看出，解析几何中空间向量所组成的线性空间是三维的。所有实系数多项式所组成的线性空间是无限维的。因为对于任意大的 N，都有 N 个线性无关的向量

$$1, \ x, \ x^2, \ \cdots, \ x^{N-1}$$

在解析几何中为了研究向量的性质，引入坐标是一个重要的步骤，对于有限维的线性空间，坐标同样是一个有力的工具。

定义 1.1.5 在 n 维线性空间 V 中，n 个线性无关的向量 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 称为 V 的一组**基**。

设 $\boldsymbol{x}$ 是 V 中任一向量，由定义 1.1.4 知，$\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n, \boldsymbol{x}$ 线性相关，由定理 1.1.1 知，$\boldsymbol{x}$ 可以由 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 唯一地线性表出，即

$$\boldsymbol{x}=a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2+\cdots+a_n\boldsymbol{e}_n \tag{1-4}$$

式(1-4)中数 $a_1, a_2, \cdots, a_n$ 称为 $\boldsymbol{x}$ 在基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 下的**坐标**，记为 $(a_1, a_2, \cdots, a_n)$。式(1-4)可以写成

$$\boldsymbol{x}=(a_1,a_2,\cdots,a_n)\begin{pmatrix}\boldsymbol{e}_1\\ \vdots\\ \boldsymbol{e}_n\end{pmatrix}=(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)\begin{pmatrix}a_1\\ \vdots\\ a_n\end{pmatrix} \tag{1-5}$$

定理 1.1.2 如果在线性空间 V 中有 n 个线性无关的向量 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，且 V 中任一向量都可以用它们线性表出，那么 V 是 n 维的，而且 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 是 V 的一组基。

〔证明〕 既然 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 线性无关，所以 V 的维数至少是 n。

设 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n, \boldsymbol{y}_{n+1}$ 是 V 中任意 $n+1$ 个向量，由定理条件它们都可以用 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 线性表出

$$y_i = c_{i1}x_1 + c_{i2}x_2 + \cdots + c_{in}x_n, \quad (i=1,2,\cdots,n) \qquad (1\text{-}6)$$

若有

$$k_1 y_1 + k_2 y_2 + \cdots + k_{n+1} y_{n+1} = 0$$

把式（1-6）代入得

$$\sum_{i=1}^{n+1} k_i c_{i1} x_1 + \sum_{i=1}^{n+1} k_i c_{i2} x_2 + \cdots + \sum_{i=1}^{n+1} k_i c_{in} x_n = 0$$

因为 x_1，x_2，$\cdots$，x_n 线性无关，故

$$\sum_{i=1}^{n+1} k_i c_{ij} = 0, \qquad (j=1,2,\cdots,n)$$

这是一个以 $n+1$ 个 k_1，k_2，$\cdots$，k_{n+1} 为未知数，由 n 个方程组成的线性齐次方程组。因此对于任何固定的数 c_{ij} 必有非零解 k_1，k_2，$\cdots$，k_{n+1}，也就是说，对于任何 y_1，y_2，$\cdots$，y_{n+1} 一定是线性相关的。由定义 1.1.4 知 V 是 n 维的，x_1，$x_2,\cdots,x_n$ 是 V 的一组基。 ▮

例 1.1.7 在线性空间 $F[x]_n$ 中

$$1,\ x,\ x^2,\ \cdots,\ x^{n-1}$$

是 n 个线性无关的向量，而且每个次数小于 n 的数域 F 上的多项式可以被它们线性表出，所以 $F[x]_n$ 是 n 维的，而 1，x，x^2，$\cdots$，x^{n-1} 是它的一组基。

在这组基下，多项式

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}$$

的坐标就是它的系数（a_0，a_1，$\cdots$，a_{n-1}）。若在 V 中取另一组基

$$1,\ (x-a),\ (x-a)^2,\ \cdots,\ (x-a)^{n-1}$$

则多项式 $f(x)$ 按泰勒公式为

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1}$$

因此在新的一组基下的坐标是

$$\left(f(a),\ f'(a),\ \cdots,\ \frac{f^{(n-1)}(a)}{(n-1)!}\right)$$

例 1.1.8 如果把复数域 c 看作是自身上的线性空间，那么它是一维的，数 1 就是一组基。如果把复数域看作是实数域上的线性空间，那么它是二维的，数 1,i 是一组基。

由此可见，维数和所考虑的数域有关。

§1.2 基变换与坐标变换

在上节例 1.1.7 中看到，同一个向量在不同基下的坐标是不同的。这节进一步研究它们

之间的关系。

设 $e_1, e_2, \cdots, e_n$ 和 $e'_1, e'_2, \cdots, e'_n$ 是 n 维线性空间 V 中两组基，它们的关系是

$$e'_1=a_{11}e_1+a_{21}e_2+\cdots+a_{n1}e_n$$

$$e'_2=a_{12}e_1+a_{22}e_2+\cdots+a_{n2}e_n$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$e'_n=a_{1n}e_1+a_{2n}e_2+\cdots+a_{nn}e_n$$

或者可以写成

$$(e'_1, e'_2, \cdots, e'_n)=(e_1, e_2, \cdots, e_n)\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{pmatrix} \tag{1-7}$$

式（1-7）中的矩阵

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \tag{1-8}$$

称为由基 $e_1, e_2, \cdots, e_n$ 到基 $e'_1, e'_2, \cdots, e'_n$ 的**过渡矩阵**，它是可逆的。

设 $x\in V$，且 x 在两组基下的坐标分别为 $(x_1,x_2,\cdots,x_n)$ 和 $(x'_1,x'_2,\cdots,x'_n)$，即

$$x=x_1e_1+x_2e_2+\cdots+x_ne_n$$

$$=x'_1e'_1+x'_2e'_2+\cdots+x'_ne'_n$$

写成矩阵形式为

$$x=(e_1, e_2, \cdots, e_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$=(e'_1, e'_2, \cdots, e'_n)\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} \tag{1-9}$$

现求 x 的坐标 x_i 与 x'_i 之间的关系。

把式（1-7）代入式（1-9）得

$$(e_1, e_2, \cdots, e_n)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (e_1, \cdots, e_n)\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}$$

由于 $e_1, e_2, \cdots, e_n$ 线性独立，故

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} \tag{1-10}$$

由于 A 是可逆的，故

$$\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}^{-1}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{1-11}$$

公式（1-10）或（1-11）称为在基变换（1-7）下，向量的坐标变换公式。

例 1.2.1 在例 1.1.7 中取 $e_1, e_2, \cdots, e_n$ 为 $1, x, x^2, \cdots, x^{n-1}$；取 $e'_1, \cdots, e'_n$ 为 $1, (x-a), (x-a)^2, \cdots, (x-a)^{n-1}$，则 $e_1, e_2, \cdots, e_n$ 到 $e'_1, \cdots, e'_n$ 的过渡矩阵 A 为

$$A = \begin{pmatrix} 1 & -a & a^2 & -a^3 & \cdots & (-a)^{n-1} \\ 0 & 1 & -2a & 3a^2 & \cdots & (n-1)(-a)^{n-2} \\ 0 & \cdots & 1 & -3a & & \cdots \\ \vdots & & & \ddots & \ddots & \\ 0 & \cdots & \cdots & \cdots & & 1 \end{pmatrix}$$

所以坐标变换公式为

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & -a & a^2 & -a^3 & \cdots & (-a)^{n-1} \\ 0 & 1 & -2a & +3a^2 & \cdots & (n-1)(-a)^{n-2} \\ 0 & 0 & 1 & -3a & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ \vdots \\ x'_n \end{pmatrix}$$

§1.3 线性子空间

1. 子空间的概念

在通常的三维几何空间中，过原点的共面向量构成一个二维向量空间，过原点的共线向量构成一个一维向量空间，而这些向量集合都是三维线性空间的子集合。在 n 维线性空间中，也可以引进相应的概念。

定义 1.3.1 n 维线性空间 V 中子集合 W 称为**线性子空间**或称为**子空间**，如果它适合

(1) 在子集合 W 中任取向量 $\boldsymbol{x}$，$\boldsymbol{y}$，则向量 $\boldsymbol{x}+\boldsymbol{y}$ 也在 W 中。

(2) 在子集合 W 中任取向量 $\boldsymbol{x}$，再任取数域 F 中一数 λ，则 $\lambda\boldsymbol{x}$ 也在 W 中。

由定义可知，在 W 中任取 m 个向量 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_m$，对于任意 m 个数 λ_1，λ_2，…，λ_m，则向量

$$\lambda_1\boldsymbol{x}_1+\lambda_2\boldsymbol{x}_2+\cdots+\lambda_m\boldsymbol{x}_m$$

也在 W 中。换句话说，子空间 W 中任意有限多个向量的任意线性组合仍在 W 中。

由定义不难证明下述定理。

定理 1.3.1 n 维线性空间 V 的子空间仍为线性空间。

证明留给读者。

既然线性子空间本身也是一个线性空间。上面引入的概念，如维数、基、坐标等，当然也可以应用到线性子空间上。因为在线性子空间中不可能比在整个空间中有更多数目的线性无关的向量。所以，任何一个线性子空间的维数不能超过整个空间的维数。

例 1.3.1 在线性空间中，由单个零向量所组成的子集合是一个线性子空间，叫做**零子空间**。

例 1.3.2 线性空间 V 本身也是 V 的一个子空间。

在线性空间中，零子空间和线性空间本身经常称为平凡子空间，而其它的线性子空间叫做非平凡子空间。

例 1.3.3 $F[x]_n$ 是线性空间 $F[x]$ 的子空间。

设 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 是线性空间 V 中一组向量，不难看出，由这组向量所有可能的线性组合

$$k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_r\boldsymbol{x}_r$$

构成的集合是非空的，而且对加法运算与数量乘法是封闭的，因而是 V 的一个子空间。这个子空间叫做由 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ **生成的子空间**，记为

$$L(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_r)$$

在有限维线性空间中，任何一个子空间都可以这样得到。事实上，设 W 是 V 的一个子空间，W 当然也是有限维的。令 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 是 W 的一组基，于是

$$W=L(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_r)$$

2. 子空间的和、交、直和

定义 1.3.2 n 维线性空间 V 的两个子空间 V_1 和 V_2 的交，是由所有既在 V_1 中，又在 V_2 中的向量构成的子集合，记作 $V_1 \cap V_2$。

显然，子空间的交仍为子空间。

定义 1.3.3 n 维线性空间 V 的两个子空间 V_1 及 V_2 的和，是由所有形如

$$z = x + y, \quad x \in V_1, \; y \in V_2$$

的向量构成的子集合，记作 $V_1 + V_2$。

显然，子空间之和仍为子空间。

例 1.3.4 在三维几何空间中，用 V_1 表示一条通过原点的直线，V_2 表示一个通过原点而且与 V_1 垂直的平面，那么 V_1 与 V_2 的交是 $\{0\}$，而 V_1 与 V_2 的和是整个空间。

例 1.3.5 用 V_1 与 V_2 分别表示齐次方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{s1}x_1 + a_{s2}x_2 + \cdots + a_{sn}x_n = 0 \end{cases}$$

与

$$\begin{cases} b_{11}x_1 + b_{12}x_2 + \cdots + b_{1n}x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ b_{t1}x_1 + b_{t2}x_2 + \cdots + b_{tn}x_n = 0 \end{cases}$$

的解空间，那么 $V_1 \cap V_2$ 就是齐次方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{s1}x_1 + a_{s2}x_2 + \cdots + a_{sn}x_n = 0 \\ b_{11}x_1 + b_{12}x_2 + \cdots + b_{1n}x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ b_{t1}x_1 + b_{t2}x_2 + \cdots + b_{tn}x_n = 0 \end{cases}$$

的解空间。

关于两个子空间的交与和的维数，有以下的定理。

定理 1.3.2（维数公式） 设 V_1 与 V_2 是线性空间 V 的两个子空间，则

$$\dim V_1 + \dim V_2 = \dim(V_1 + V_2) + \dim(V_1 \cap V_2) \qquad (1\text{-}12)$$

〔证明〕 设 $\dim V_1 = n_1$，$\dim V_2 = n_2$，$\dim(V_1 \cap V_2) = m$，取 $V_1 \cap V_2$ 的一组基

$$x_1, x_2, \cdots, x_m$$

它可以扩充成 V_1 的一组基

$$x_1, x_2, \cdots, x_m, y_1, \cdots, y_{n_1-m}$$

也可以扩充成 V_2 的一组基

$$x_1, x_2, \cdots, x_m, z_1, \cdots, z_{n_2-m}$$

此即

$$V_1 = L(x_1, \cdots, x_m, y_1, \cdots, y_{n_1-m})$$

$$V_2 = L(x_1, \cdots, x_m, z_1, \cdots, z_{n_2-m})$$

所以

$$V_1 + V_2 = L(x_1, \cdots, x_m, y_1, \cdots, y_{n_1-m}, z_1, \cdots, z_{n_2-m})$$

设

$$k_1x_1 + \cdots + k_mx_m + p_1y_1 + \cdots + p_{n_1-m}y_{n_1-m} + q_1z_1 + \cdots + q_{n_2-m}z_{n_2-m} = 0$$

命

$$x = k_1x_1 + \cdots + k_mx_m + p_1y_1 + \cdots + p_{n_1-m}y_{n_1-m} = -q_1z_1 - \cdots - q_{n_2-m}z_{n_2-m}$$

由第一个等式知 $x \in V_1$，由第二个等式知 $x \in V_2$，于是 $x \in V_1 \cap V_2$，故可令

$$x = l_1x_1 + \cdots + l_mx_m$$

则

$$l_1x_1 + \cdots + l_mx_m = -q_1z_1 - \cdots - q_{n_2-m}z_{n_2-m}$$

即

$$l_1x_1 + \cdots + l_mx_m + q_1z_1 + \cdots + q_{n_2-m}z_{n_2-m} = 0$$

由于 $x_1, x_2, \cdots, x_m, z_1, \cdots, z_{n_2-m}$ 线性无关，所以

$$l_1 = \cdots = l_m = q_1 = \cdots = q_{n_2-m} = 0$$

因而 $x = 0$ ，从而有

$$k_1x_1 + \cdots + k_mx_m + p_1y_1 + \cdots + p_{n_1-m}y_{n_1-m} = 0$$

由于 $x_1, \cdots, x_m, y_1, \cdots, y_{n_1-m}$ 线性无关，又得

$$k_1 = \cdots = k_m = p_1 = \cdots = p_{n_1-m} = 0$$

这就证明了 $x_1, \cdots, x_m, y_1, \cdots, y_{n_1-m}, z_1, \cdots, z_{n_2-m}$ 线性无关，因而它是 V_1+V_2 的一组基，V_1+V_2 的维数为 n_1+n_2-m。于是维数公式成立。 ▌

子空间的直和是子空间和的一个重要特殊情况。

定义 1.3.4 设 V_1，V_2 是线性空间 V 的子空间，如果 V_1+V_2 中每个向量的分解式

$$x = x_1 + x_2, \quad x_1 \in V_1, \ x_2 \in V_2$$

是唯一的，这个和就称为直和，记为 $V_1 \dotplus V_2$。

定理 1.3.3 和 $V_1 + V_2$ 是直和的充要条件是等式

$$x_1 + x_2 = 0, \quad x_i \in V_i \quad (i = 1, 2)$$

只在 x_i 全为零向量时成立。

〔证明〕 必要性 由直和定义，$V_1 \dotplus V_2$ 中每个向量分解式是唯一的，所以零向量的分解也是唯一的。

充分性 设 $x \in V_1 + V_2$ 有两个分解式

$$x = x_1 + x_2 = y_1 + y_2, \quad x_i, \ y_i \in V_i \ (i = 1, 2)$$

于是

$$(x_1 - y_1) + (x_2 - y_2) = 0$$

其中 $x_i - y_i \in V_i$ ($i = 1, 2$)，由定理条件有

$$x_i - y_i = 0 \ , \ x_i = y_i \ (i = 1, 2)$$

此即向量 x 的分解式是唯一的。 ▮

推论 1.3.1 和 $V_1 + V_2$ 为直和的充要条件是

$$V_1 \cap V_2 = \{0\}$$

〔证明〕 充分性 设

$$x_1 + x_2 = 0, \quad x_i \in V_i \ (i = 1, 2)$$

那么

$$x_1 = -x_2 \in V_1 \cap V_2$$

由假设知

$$x_1 = x_2 = 0$$

此即 $V_1 + V_2$ 是直和。

必要性 任取向量 $x \in V_1 \cap V_2$，于是零向量可以写成

$$x + (-x) = 0 \ , \ x \in V_1, \ (-x) \in V_2$$

因为是直和，所以

$$x = -x = 0$$

此即

$$V_1 \cap V_2 = \{0\}$$ ▮

定理 1.3.4 和 $V_1 + V_2$ 是直和的充要条件是在 V_1 中任取一组基：$x_{11}, x_{12}, \cdots, x_{1n_1}$ 与在 V_2 中任取一组基：$x_{21}, x_{22}, \cdots, x_{2n_2}$，则向量组

$$x_{11}, x_{12}, \cdots, x_{1n_1}, x_{21}, x_{22}, \cdots, x_{2n_2}$$

构成 V_1+V_2 的一组基。

〔证明〕 必要性 设 $x \in V_1+V_2$，则

$$x=x_1+x_2, \qquad x_i \in V_i, \ (i=1, 2)$$

而

$$x_1=\sum_{j=1}^{n_1} k_{1j} x_{1j}, \quad x_2=\sum_{j=1}^{n_2} k_{2j} x_{2j}$$

故得

$$x=\sum_{j=1}^{n_1} k_{1j} x_{1j}+\sum_{j=1}^{n_2} k_{2j} x_{2j}$$

此即 x 可由 x_{1j}，x_{2k}（$j=1, 2, \cdots, n_1$；$k=1, 2, \cdots, n_2$）线性表出。现证 x_{1j}，x_{2k} 线性无关。事实上，设

$$\sum_{j=1}^{n_1} c_{1j} x_{1j}+\sum_{k=1}^{n_2} c_{2k} x_{2k}=0$$

由定理 1.3.3 知

$$\sum_{j=1}^{n_1} c_{1j} x_{1j}=0, \quad \sum_{k=1}^{n_2} c_{2k} x_{2k}=0$$

因为 x_{1j} 线性无关，故 $c_{1j}=0$（$j=1, 2, \cdots, n_1$）。又因 x_{2k} 线性无关，故 $c_{2k}=0$（$k=1, 2, \cdots, n_2$）。于是 x_{1j}，x_{2k} 线性无关，从而 x_{1j}，x_{2k} 构成 V_1+V_2 的一组基。

充分性 设 $x \in V_1+V_2$ 有两个分解式

$$x=x_1+x_2=y_1+y_2, \quad x_i, \ y_i \in V_i \ (i=1, 2)$$

则

$$(x_1-y_1)+(x_2-y_2)=0$$

命

$$z_1=x_1-y_1, \quad z_2=x_2-y_2$$

显然，$z_1 \in V_1$，$z_2 \in V_2$，因此

$$z_1=\sum_{j=1}^{n_1} c_{1j} x_{1j}, \quad z_2=\sum_{k=1}^{n_2} c_{2k} x_{2k}$$

于是

$$z_1+z_2=\sum_{j=1}^{n_1} c_{1j} x_{1j}+\sum_{k=1}^{n_2} c_{2k} x_{2k}=0$$

由于 x_{1j}，x_{2k} 构成 V_1+V_2 的一组基，所以 $c_{1j}=c_{2k}=0$（$j=1, 2, \cdots, n_1$；$k=1, 2, \cdots, n_2$）。此即

$$\boldsymbol{z}_1=\boldsymbol{z}_2=\boldsymbol{0}$$

所以 $$\boldsymbol{x}_1=\boldsymbol{y}_1,\ \boldsymbol{x}_2=\boldsymbol{y}_2$$

这就证明了 V_1+V_2 是直和。 ∎

定理 1.3.5 和 V_1+V_2 是直和的充要条件为

$$\dim(V_1+V_2)=\dim V_1+\dim V_2 \tag{1-13}$$

证明作为练习留给读者。

定理 1.3.6 设 U 是 V 的一个子空间，则一定存在一个子空间 W，使得

$$V=U\dot{+}W$$

〔证明〕 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ 是 U 的一组基，把它扩充为 V 的一组基 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m, \boldsymbol{x}_{m+1}, \cdots, \boldsymbol{x}_n$。令

$$W=L(\boldsymbol{x}_{m+1}, \boldsymbol{x}_{m+2}, \cdots, \boldsymbol{x}_n)$$

则 W 即为所需。 ∎

§1.4 线性变换

1. 线性变换的概念

定义 1.4.1 设 V_1 与 V_2 分别表示数域 F 上的两个线性空间，映射 $\mathscr{A}: V_1 \to V_2$ 叫做**线性变换**，如果对于 V_1 的任何两个矢量 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in V_1$ 和任何数 $\lambda \in F$，有

$$\mathscr{A}(\boldsymbol{x}_1+\boldsymbol{x}_2)=\mathscr{A}(\boldsymbol{x}_1)+\mathscr{A}(\boldsymbol{x}_2)$$

$$\mathscr{A}(\lambda\boldsymbol{x}_1)=\lambda\mathscr{A}(\boldsymbol{x}_1)$$

即映射 $\mathscr{A}$ 是可加的和齐次的。

例 1.4.1 设 $A=(a_{ij})$ 是实 $m\times n$ 阶矩阵，映射 $\mathscr{A}: R^n \to R^m$ 由下式确定

$$\mathscr{A}(\boldsymbol{x})=A\boldsymbol{x},\ \forall \boldsymbol{x}\in R^n$$

不难验证 $\mathscr{A}$ 是线性变换。

设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r \in V_1$，$\lambda_1, \lambda_2, \cdots, \lambda_r \in P$，由线性变换 $\mathscr{A}$ 的定义可得

$$\begin{aligned}&\mathscr{A}(\lambda_1\boldsymbol{x}_1+\cdots+\lambda_r\boldsymbol{x}_r)\\ &\quad=\lambda_1\mathscr{A}(\boldsymbol{x}_1)+\cdots+\lambda_r\mathscr{A}(\boldsymbol{x}_r)\end{aligned}$$

特别地

$$\mathscr{A}(\boldsymbol{0})=\boldsymbol{0}$$

$$\mathscr{A}(-\boldsymbol{x})=-\mathscr{A}(\boldsymbol{x}),\qquad \boldsymbol{x}\in V_1$$

$$\mathscr{A}(\boldsymbol{x}_1-\boldsymbol{x}_2)=\mathscr{A}(\boldsymbol{x}_1)-\mathscr{A}(\boldsymbol{x}_2)$$

2. 线性变换与矩阵

设 $x_1, x_2, \cdots, x_n$ 是 V_1 的一组基，$y_1, y_2, \cdots, y_m$ 是 V_2 的一组基。因为 $\mathscr{A}(x_j) \in V_2$，故

$$\mathscr{A}(x_j) = \sum_{i=1}^{m} a_{ij} y_i \quad (j=1, 2, \cdots, n)$$

或写成

$$\begin{aligned}
&\mathscr{A}(x_1, x_2, \cdots, , x_n) \\
&= (\mathscr{A}(x_1), \mathscr{A}(x_2), \cdots, \mathscr{A}(x_n)) \\
&= \left(\sum_{i=1}^{m} a_{i1} y_i, \sum_{i=1}^{m} a_{i2} y_i, \cdots, \sum_{i=1}^{m} a_{in} y_i\right) \\
&= (y_1, y_2, \cdots, y_m) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}
\end{aligned} \tag{1-14}$$

命

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{1-15}$$

所以式（1-14）可以写为

$$\mathscr{A}(x_1, x_2, \cdots, x_n) = (y_1, y_2, \cdots, y_m) A \tag{1-16}$$

矩阵 A 称为线性变换 $\mathscr{A}$ 在基（$x_1, x_2, \cdots, x_n$）与（$y_1, y_2, \cdots, y_m$）下的矩阵表示。

设 $x \in V_1$，故

$$x = \sum_{i=1}^{n} x_i x_i$$

它的象 $\mathscr{A}(x) \in V_2$ 可写为

$$\mathscr{A}(x) = \sum_{j=1}^{m} y_j y_j$$

另一方面

$$\mathscr{A}(x) = \mathscr{A}\left(\sum_{i=1}^{n} x_i x_i\right) = \sum_{i=1}^{n} x_i \mathscr{A}(x_i)$$

$$=\sum_{i=1}^{n}x_i\sum_{j=1}^{m}a_{ji}y_j=\sum_{j=1}^{m}\left(\sum_{i=1}^{n}a_{ji}x_i\right)y$$

因为 $\mathscr{A}(\boldsymbol{x})$ 的坐标是唯一的，故

$$y_j=\sum_{i=1}^{n}a_{ji}x_i\quad(j=1,\ 2,\ \cdots,\ m)$$

写成矩阵形式为

$$\begin{pmatrix}y_1\\y_2\\\vdots\\y_m\end{pmatrix}=\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\cdots&\cdots&\cdots&\cdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\tag{1-17}$$

式（1-17）称为线性变换在给定基（$\boldsymbol{x}_1,\ \boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$）与（$\boldsymbol{y}_1,\ \boldsymbol{y}_2,\ \cdots,\ \boldsymbol{y}_m$）下向量坐标变换公式。

特别地，当 $V_1=V_2=V$ 时，取一组基为 $\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_n$，则

$$\mathscr{A}(\boldsymbol{x}_j)=\sum_{i=1}^{n}a_{ij}\boldsymbol{x}_i$$

于是

$$\begin{aligned}&\mathscr{A}(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_n)\\=&(\mathscr{A}\boldsymbol{x}_1,\ \mathscr{A}\boldsymbol{x}_2,\ \cdots,\ \mathscr{A}\boldsymbol{x}_n)\\=&(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_n)\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\cdots&\cdots&\cdots&\cdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{pmatrix}\end{aligned}\tag{1-18}$$

线性变换 $\mathscr{A}$ 在基 $\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_n$ 下的矩阵表示

$$A=\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\cdots&\cdots&\cdots&\cdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{pmatrix}\tag{1-19}$$

是 n 阶方阵。向量坐标变换公式为

$$\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}=\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\cdots&\cdots&\cdots&\cdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\tag{1-20}$$

下面研究的线性变换$\mathscr{A}$及其矩阵A均属$V_1=V_2=V$的情况。

在n维线性空间V中取定一组基$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_n$，任取n阶矩阵$A=(a_{ij})$，则可构造n个向量

$$\boldsymbol{y}_i=\sum_{j=1}^{n}a_{ij}\boldsymbol{x}_j \quad (i=1,\ 2,\ \cdots,\ n) \qquad (1\text{-}21)$$

反之，任取n个向量$\boldsymbol{y}_1$，$\boldsymbol{y}_2$，…，$\boldsymbol{y}_n\in V$，由式(1-21)可以唯一确定一个n阶矩阵A。这样，n个n维向量和n阶矩阵一一对应。

定理 1.4.1 在n维线性空间V中取定一组基$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_n$，对V中任意n个向量$\boldsymbol{y}_1$，$\boldsymbol{y}_2$，…，$\boldsymbol{y}_n$，存在一个且只有一个线性变换$\mathscr{A}$，它把$\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$分别映为$\boldsymbol{y}_1$，$\boldsymbol{y}_2$，…，$\boldsymbol{y}_n$，即

$$\boldsymbol{y}_j=\mathscr{A}(\boldsymbol{x}_j) \quad (j=1,\ 2,\ \cdots,\ n)$$

〔证明〕 在V中任取向量$\boldsymbol{x}=\sum_{j=1}^{n}x_j\boldsymbol{x}_j$，定义映射$\mathscr{A}$

$$\boldsymbol{x}=\sum_{j=1}^{n}x_j\boldsymbol{x}_j\longrightarrow\sum_{j=1}^{n}x_j\boldsymbol{y}_j=\boldsymbol{y}$$

则$\mathscr{A}$是V到V的一个线性变换，事实上，

$$\begin{aligned}&\mathscr{A}(\lambda\boldsymbol{x}+\mu\boldsymbol{x}')\\&=\mathscr{A}\Big(\lambda\sum_{j=1}^{n}x_j\boldsymbol{x}_j+\mu\sum_{j=1}^{n}x'_j\boldsymbol{x}_j\Big)\\&=\mathscr{A}\Big(\sum_{j=1}^{n}(\lambda x_j+\mu x'_j)\boldsymbol{x}_j\Big)\\&=\sum_{j=1}^{n}(\lambda x_j+\mu x'_j)\boldsymbol{y}_j\\&=\sum_{j=1}^{n}\lambda x_j\boldsymbol{y}_j+\sum_{j=1}^{n}\mu x'_j\boldsymbol{y}_j\\&=\lambda\boldsymbol{y}+\mu\boldsymbol{y}'=\lambda\mathscr{A}(\boldsymbol{x})+\mu\mathscr{A}(\boldsymbol{x}')\end{aligned}$$

且

$$\mathscr{A}(\boldsymbol{x}_j)=\boldsymbol{y}_j \quad (j=1,\ 2,\ \cdots,\ n)$$

现证唯一性。如果存在线性变换$\mathscr{B}$，使得

$$\mathscr{B}(\boldsymbol{x}_j)=\boldsymbol{y}_j, \quad (j=1,\ 2,\ \cdots,\ n)$$

那么对V中的任意向量$\boldsymbol{x}=\sum_{j=1}^{n}x_j\boldsymbol{x}_j$，有

$$\mathscr{B}(\boldsymbol{x})=\mathscr{B}\Big(\sum_{j=1}^{n}x_j\boldsymbol{x}_j\Big)=\sum_{j=1}^{n}x_j\,\mathscr{B}(\boldsymbol{x}_j)$$

$$=\sum_{j=1}^{n}\tau_j y_j = \mathscr{A}(x)$$

即唯一性成立。]

定理 1.4.1 表明，n 维线性空间中的线性变换在固定基下和 n 阶矩阵一一对应。

显然，同一个线性变换在不同基下的矩阵表示是不同的，但它们之间却有一个重要的关系。

定理 1.4.2 在线性空间 V 中取定基 $x_1, x_2, \cdots, x_n$ 与基 $y_1, y_2, \cdots, y_n$，假设基变换为

$$(y_1, \cdots, y_n)=(x_1, \cdots, x_n)P$$

线性变换 $\mathscr{A}$ 在 $x_1, \cdots, x_n$ 下的矩阵表示为 A，在 $y_1, \cdots, y_n$ 下的矩阵表示为 B，则

$$B=P^{-1}AP$$

〔证明〕 由定理条件知有

$$\mathscr{A}(x_1, \cdots, x_n)=(x_1, \cdots, x_n)A$$

$$\mathscr{A}(y_1, \cdots, y_n)=(y_1, \cdots, y_n)B$$

$$(y_1, \cdots, y_n)=(x_1, \cdots, x_n)P$$

故得

$$\mathscr{A}(y_1, \cdots, y_n)=\mathscr{A}(x_1, \cdots, x_n)P$$
$$=(x_1, \cdots, x_n)AP$$

和

$$\mathscr{A}(y_1, \cdots, y_n)=(y_1, \cdots, y_n)B$$
$$=(x_1, \cdots, x_n)PB$$

因此

$$(x_1, \cdots, x_n)AP=(x_1, \cdots, x_n)PB$$

根据 $x_1, \cdots, x_n$ 线性无关就有

$$AP=PB$$

或

$$B=P^{-1}AP$$]

定义 1.4.2 n 阶矩阵 A 和 B 称为相似的，如果存在 n 阶满秩矩阵 P，使得

$$B=P^{-1}AP \qquad (1\text{-}22)$$

不难证明，相似关系有下列性质：

（1） 反身性 A 和 A 相似；

（2） 对称性 若 A 和 B 相似，则 B 和 A 相似；

（3） 传递性 若 A 和 B 相似，B 和 C 相似，则 A 和 C 相似。

因此相似是一个关系。经常用 $A\sim B$ 表示 A 和 B 相似。

例 1.4.2 设线性变换 $\mathscr{A}$ 在基 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，$\boldsymbol{x}_3$，$\boldsymbol{x}_4$ 下的矩阵表示为

$$\begin{pmatrix}1 & 3 & 2 & 8\\5 & 7 & 0 & 1\\3 & 0 & 1 & 3\\1 & 1 & 2 & 2\end{pmatrix}$$

试求线性变换 $\mathscr{A}$ 在基 $\boldsymbol{y}_1=\boldsymbol{x}_1$，$\boldsymbol{y}_2=\boldsymbol{x}_1+\boldsymbol{x}_2$，$\boldsymbol{y}_3=\boldsymbol{x}_1+\boldsymbol{x}_2+\boldsymbol{x}_3$，$\boldsymbol{y}_4=\boldsymbol{x}_1+\boldsymbol{x}_2+\boldsymbol{x}_3+\boldsymbol{x}_4$ 下的矩阵表示。

〔解〕 法一 由矩阵 A 的定义知

$$\mathscr{A}(\boldsymbol{x}_1)=\boldsymbol{x}_1+5\boldsymbol{x}_2+3\boldsymbol{x}_3+\boldsymbol{x}_4$$

$$\mathscr{A}(\boldsymbol{x}_2)=3\boldsymbol{x}_1+7\boldsymbol{x}_2\qquad+\boldsymbol{x}_4$$

$$\mathscr{A}(\boldsymbol{x}_3)=2\boldsymbol{x}_1\qquad+\boldsymbol{x}_3+2\boldsymbol{x}_4$$

$$\mathscr{A}(\boldsymbol{x}_4)=8\boldsymbol{x}_1+\boldsymbol{x}_2+3\boldsymbol{x}_3+2\boldsymbol{x}_4$$

于是

$$\mathscr{A}(\boldsymbol{y}_1)=\mathscr{A}(\boldsymbol{x}_1)=\boldsymbol{x}_1+5\boldsymbol{x}_2+3\boldsymbol{x}_3+\boldsymbol{x}_4$$

$$=-4\boldsymbol{y}_1+2\boldsymbol{y}_2+2\boldsymbol{y}_3+\boldsymbol{y}_4$$

$$\mathscr{A}(\boldsymbol{y}_2)=\mathscr{A}(\boldsymbol{x}_1+\boldsymbol{x}_2)=4\boldsymbol{x}_1+12\boldsymbol{x}_2+3\boldsymbol{x}_3+2\boldsymbol{x}_4$$

$$=-8\boldsymbol{y}_1+9\boldsymbol{y}_2+\boldsymbol{y}_3+2\boldsymbol{y}_4$$

$$\mathscr{A}(\boldsymbol{y}_3)=\mathscr{A}(\boldsymbol{x}_1+\boldsymbol{x}_2+\boldsymbol{x}_3)$$

$$=6\boldsymbol{x}_1+12\boldsymbol{x}_2+4\boldsymbol{x}_3+4\boldsymbol{x}_4$$

$$=-6\boldsymbol{y}_1+8\boldsymbol{y}_2+4\boldsymbol{y}_4$$

$$\mathscr{A}(\boldsymbol{y}_4)=\mathscr{A}(\boldsymbol{x}_1+\boldsymbol{x}_2+\boldsymbol{x}_3+\boldsymbol{x}_4)$$

$$=14\boldsymbol{x}_1+13\boldsymbol{x}_2+7\boldsymbol{x}_3+6\boldsymbol{x}_4$$

$$=\boldsymbol{y}_1+6\boldsymbol{y}_2+\boldsymbol{y}_3+6\boldsymbol{y}_4$$

所以在新的基 $\boldsymbol{y}_1$，$\boldsymbol{y}_2$，$\boldsymbol{y}_3$，$\boldsymbol{y}_4$ 下 $\mathscr{A}$ 的矩阵表示为

$$B=\begin{pmatrix}-4 & -8 & -6 & 1\\2 & 9 & 8 & 6\\2 & 1 & 0 & 1\\1 & 2 & 4 & 6\end{pmatrix}$$

法二 可以用定理 1.4.2 求 B。由基 $\boldsymbol{x}_i$ 到 $\boldsymbol{y}_i$ 的过渡矩阵

$$P=\begin{pmatrix}1&1&1&1\\0&1&1&1\\0&0&1&1\\0&0&0&1\end{pmatrix}$$

从而

$$P^{-1}=\begin{pmatrix}1&-1&0&0\\0&1&-1&0\\0&0&1&-1\\0&0&0&1\end{pmatrix}$$

所以

$$P^{-1}AP=\begin{pmatrix}1&-1&0&0\\0&1&-1&0\\0&0&1&-1\\0&0&0&1\end{pmatrix}\begin{pmatrix}1&3&2&8\\5&7&0&1\\3&0&1&3\\1&1&8&2\end{pmatrix}\begin{pmatrix}1&1&1&1\\0&1&1&1\\0&0&1&1\\0&0&0&1\end{pmatrix}$$

$$=\begin{pmatrix}-4&-8&-6&1\\2&9&8&6\\2&1&0&1\\1&2&4&6\end{pmatrix}=B$$

3. 线性变换的值域与核

定义 1.4.3 设$\mathscr{A}$是线性空间 V 的一个线性变换，线性空间 V 中所有象向量构成的集合称为线性变换$\mathscr{A}$的值域，记为$\mathscr{A}V$。即

$$\mathscr{A}V=\{y=\mathscr{A}(x)\mid x\in V\}$$

显然$\mathscr{A}V$ 是 V 的子空间。

定理 1.4.3 设$\mathscr{A}$是 n 维线性空间 V 的一个线性变换，x_1，x_2，…，x_n 是 V 的一组基，$\mathscr{A}$ 在这组基下的矩阵表示为 A，则

（1） $\mathscr{A}V=L(\mathscr{A}(x_1),\ \mathscr{A}(x_2),\ \cdots,\ \mathscr{A}(x_n))$

（2） $\mathscr{A}V$ 的秩$=A$ 的秩

〔证明〕 第一部分结论的证明留给读者。现证第二部分结论。

因为值域 $\mathscr{A}V$ 是由 n 个向量

$$y_j=\mathscr{A}x_j=a_{j1}x_1+a_{j2}x_2+\cdots+a_{jn}x_n,\quad (j=1,\ 2,\ \cdots,\ n) \tag{1-23}$$

生成的子空间。所以向量组 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n$ 中任一极大线性无关组 $\boldsymbol{y}_{j1}, \boldsymbol{y}_{j2}, \cdots, \boldsymbol{y}_{jr}$ 是 $\mathscr{A}V$ 的一组基。另一方面，V 中每一向量 $\boldsymbol{y}=\sum_{i=1}^{n}x_i\boldsymbol{x}_i$ 唯一地与 R^n 中一向量 $(x_1, x_2, \cdots, x_n)$ 一一对应，且在此对应下线性关系保持不变。因此在 R^n 中有 n 个向量 $\boldsymbol{y}'_1, \boldsymbol{y}'_2, \cdots, \boldsymbol{y}'_n$，其中

$$\boldsymbol{y}_j \leftrightarrow \boldsymbol{y}'_j=(a_{j1}, a_{j2}, \cdots, a_{jn}) \quad (j=1, 2, \cdots, n)$$

于是 $\boldsymbol{y}'_{j1}, \boldsymbol{y}'_{j2}, \cdots, \boldsymbol{y}'_{jr}$ 是向量组 $\boldsymbol{y}'_1, \boldsymbol{y}'_2, \cdots, \boldsymbol{y}'_n$ 中极大线性无关向量组。

由式（1-23）知

$$A=\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}$$

因此 $\boldsymbol{y}'_1, \boldsymbol{y}'_2, \cdots, \boldsymbol{y}'_n$ 是矩阵 A 的 n 个列向量。由于矩阵 A 的秩等于列向量中极大线性无关向量组的向量个数，因此也等于 $\boldsymbol{y}'_1, \cdots, \boldsymbol{y}'_n$ 中极大线性无关向量组的向量个数，即 $\mathscr{A}V$ 的维数。 】

定义 1.4.4 值域 $\mathscr{A}V$ 的维数称为线性变换 $\mathscr{A}$ 的秩。

显然，线性变换 $\mathscr{A}$ 的秩等于 $\mathscr{A}$ 在任何一组基下矩阵表示 A 的秩。

定义 1.4.5 设 $\mathscr{A}$ 是线性空间 V 的一个线性变换，线性空间 V 中，所有被 $\mathscr{A}$ 变成零向量的向量组成的集合称为 $\mathscr{A}$ 的核，用 $\mathscr{A}^{-1}(0)$ 表示。即

$$\mathscr{A}^{-1}(0)=\{\boldsymbol{x}\in V \mid \mathscr{A}(\boldsymbol{x})=0\}$$

显然 $\mathscr{A}^{-1}(0)$ 是 V 的子空间。$\mathscr{A}^{-1}(0)$ 的维数称为 $\mathscr{A}$ 的零度。

定理 1.4.4 设 $\mathscr{A}$ 是 n 维线性空间的线性变换，则

$$\mathscr{A}\text{的秩}+\mathscr{A}\text{的零度}=n$$

〔证明〕 设 $\mathscr{A}$ 的零度为 r，在核 $\mathscr{A}^{-1}(0)$ 中取一组基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$，现把它扩充成 V 的一组基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r, \boldsymbol{x}_{r+1}, \cdots, \boldsymbol{x}_n$。根据定理 1.4.3，值域 $\mathscr{A}V$ 是由

$$\mathscr{A}\boldsymbol{x}_1, \mathscr{A}\boldsymbol{x}_2, \cdots, \mathscr{A}\boldsymbol{x}_r, \mathscr{A}\boldsymbol{x}_{r+1}, \cdots, \mathscr{A}\boldsymbol{x}_n$$

线性生成的，但是 $\mathscr{A}\boldsymbol{x}_1=\mathscr{A}\boldsymbol{x}_2=\cdots=\mathscr{A}\boldsymbol{x}_r=0$，所以 $\mathscr{A}V$ 是由

$$\mathscr{A}\boldsymbol{x}_{r+1}, \mathscr{A}\boldsymbol{x}_{r+2}, \cdots, \mathscr{A}\boldsymbol{x}_n$$

线性生成。现在来证明它们是线性无关的。

设

$$\sum_{i=r+1}^{n}k_i\mathscr{A}\boldsymbol{x}_i=0$$

根据 $\mathscr{A}$ 的线性性便得

$$\mathscr{A}\left(\sum_{i=r+1}^{n} k_i \boldsymbol{x}_i\right)=0$$

这说明 $\sum_{i=r+1}^{n} k_i \boldsymbol{x}_i \in \mathscr{A}^{-1}(0)$，因此 $\sum_{i=r+1}^{n} k_i \boldsymbol{x}_i$ 可由 $\boldsymbol{x}_1$，…，$\boldsymbol{x}_r$ 线性表出，即

$$\sum_{i=r+1}^{n} k_i \boldsymbol{x}_i=\sum_{i=1}^{r} c_i \boldsymbol{x}_i$$

从 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_n$ 线性无关推出

$$k_i=0, \quad c_j=0 \quad (i=1,\ 2,\ \cdots,\ r;\ j=r+1,\ \cdots,\ n)$$

因此 $\mathscr{A}\boldsymbol{x}_{r+1}$，$\mathscr{A}\boldsymbol{x}_{r+2}$，…，$\mathscr{A}\boldsymbol{x}_n$ 线性无关，它是 $\mathscr{A}V$ 的一组基，$\mathscr{A}$ 的秩为 $n-r$，于是

$$\mathscr{A}\text{的秩}+\mathscr{A}\text{的零度}=n$$]

应该指出，虽然子空间 $\mathscr{A}V$ 与 $\mathscr{A}^{-1}(0)$ 的维数之和为 n，但是 $\mathscr{A}V+\mathscr{A}^{-1}(0)$ 并不一定是整个空间。例如，在线性空间 $F[x]_n$ 中，令

$$\mathscr{D}(f(x))=f'(x)$$

容易验证 $\mathscr{D}$ 是 $F[x]_n$ 的线性变换，$\mathscr{D}$ 的值域是 $F[x]_{n-1}$，$\mathscr{D}$ 的核是数域 F，显然 $F[x]_{n-1}+F\neq F[x]_n$。

4. 线性变换的不变子空间

下面介绍不变子空间的概念，利用它讨论线性变换和子空间的关系，由此可以进一步揭示线性变换的矩阵与线性变换的内在联系。

定义 1.4.6 设 $\mathscr{A}$ 是数域 F 上线性空间 V 的线性变换，W 是 V 的子空间，若 $\mathscr{A}W\subset W$，则称 W 是 $\mathscr{A}$ 的**不变子空间**。

例如，V 本身就是不变子空间，它是任何线性变换 $\mathscr{A}$ 的不变子空间。线性变换 $\mathscr{A}$ 的值域 $\mathscr{A}V$ 与核 $\mathscr{A}^{-1}(0)$ 都是 $\mathscr{A}$ 的不变子空间。

设 W 是 $\mathscr{A}$ 的不变子空间，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 是 W 的一组基，再在线性空间 V 中补足 $n-r$ 个向量，使得 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$，$\boldsymbol{x}_{r+1}$，…，$\boldsymbol{x}_n$ 是 V 的一组基，现在来求 $\mathscr{A}$ 在这组基下的矩阵表示。

因为，$\mathscr{A}W\subset W$，所以 $\mathscr{A}(\boldsymbol{x}_1)$，$\mathscr{A}(\boldsymbol{x}_2)$，…，$\mathscr{A}(\boldsymbol{x}_r)$ 都是 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$ 的线性组合，因此

$$\mathscr{A}(\boldsymbol{x}_1)=a_{11}\boldsymbol{x}_1+a_{12}\boldsymbol{x}_2+\cdots+a_{1r}\boldsymbol{x}_r$$

$$\mathscr{A}(\boldsymbol{x}_2)=a_{21}\boldsymbol{x}_1+\cdots+a_{2r}\boldsymbol{x}_r$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\mathscr{A}(\boldsymbol{x}_r)=a_{r1}\boldsymbol{x}_1+\cdots+a_{rr}\boldsymbol{x}_r$$

$$\mathscr{A}(\boldsymbol{x}_{r+1})=a_{r+1,1}\boldsymbol{x}_1+\cdots+a_{r+1,r}\boldsymbol{x}_r+a_{r+1,r+1}\boldsymbol{x}_{r+1}+\cdots+a_{r+1,n}\boldsymbol{x}_n$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\mathscr{A}(x_n)=a_{n1}x_1+\cdots+a_{nr}x_r+a_{n,r+1}x_{r+1}+\cdots+a_{nn}x_n$$

所以对这组基而言，线性变换$\mathscr{A}$的矩阵表示为

$$A=\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{r1} & a_{r+1,1} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{r2} & a_{r+1,2} & \cdots & a_{n2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{1r} & a_{2r} & \cdots & a_{rr} & a_{r+1,r} & \cdots & a_{nr} \\ & & & & a_{r+1,r+1} & \cdots & a_{n,r+1} \\ & 0 & & & \cdots & \cdots & \cdots \\ & & & & a_{r+1,n} & \cdots & a_{nn} \end{pmatrix}$$

$$=\begin{pmatrix} A_1 & A_3 \\ 0 & A_2 \end{pmatrix} \qquad (1\text{-}24)$$

这是一个准上三角形矩阵。

反之，若$\mathscr{A}$在x_1，…，x_r，x_{r+1}，…，x_n下的矩阵表示为式（1-24），容易验证，由x_1，…，x_r生成的子空间是$\mathscr{A}$的不变子空间。

上面讨论了线性变换的矩阵表示是准三角形和不变子空间的关系，下面讨论矩阵表示是准对角形和不变子空间的关系。

设V是线性变换$\mathscr{A}$的若干个不变子空间的直和

$$V=W_1\dot{+}W_2+\cdots\dot{+}W_s$$

在每一个子空间W_i中取基

$$x_{i1},\ x_{i2},\ \cdots,\ x_{ir_i}\quad (i=1,\ 2,\ \cdots,\ s) \qquad (1\text{-}25)$$

其中$r_1+r_2+\cdots+r_s=n=\dim V$。把这些基合起来构成$V$的一个基。容易验证，在这组基下$\mathscr{A}$的矩阵表示

$$A=\begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_s \end{pmatrix} \qquad (1\text{-}26)$$

其中A_i是$r_i\times r_i$阶矩阵。

反之，若线性变换$\mathscr{A}$在基（1-25）下的矩阵表示是式（1-26），则由式（1-25）生成的子空间W_i是$\mathscr{A}$的不变子空间。

这个证明作为练习留给读者。

§1.5　特征值与特征向量

1. 特征值和特征向量

这一节我们来讨论在矩阵理论中具有重要作用的特征值和特征向量的概念，在后面的讨论中将会看到，特征值和特征向量在化矩阵为对角形的问题中，占有特殊的地位。

定义 1.5.1 设$\mathscr{A}$是线性空间V的一个线性变换，如果对于$\lambda_0\in F$，存在一个非零向量$\boldsymbol{x}$，使得

$$\mathscr{A}\boldsymbol{x}=\lambda_0\boldsymbol{x} \tag{1-27}$$

成立，则称λ_0为$\mathscr{A}$的**特征值**，$\boldsymbol{x}$称为$\mathscr{A}$的属于λ_0的**特征向量**。

下面讨论特征值和特征向量的求法。

设V是数域F上的n维线性空间，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_n$是V的一组基，$\mathscr{A}$是V的一个线性变换，$\mathscr{A}$在这组基下的矩阵为A，λ_0为$\mathscr{A}$的特征值，$\boldsymbol{x}$为$\mathscr{A}$的属于λ_0的特征向量。

设

$$\boldsymbol{x}=x_1^{(0)}\boldsymbol{x}_1+x_2^{(0)}\boldsymbol{x}_2+\cdots+x_n^{(0)}\boldsymbol{x}_n$$

由式（1-20）把式（1-27）改写为

$$A\begin{pmatrix}x_1^{(0)}\\ \vdots\\ x_n^{(0)}\end{pmatrix}=\lambda_0\begin{pmatrix}x_1^{(0)}\\ \vdots\\ x_n^{(0)}\end{pmatrix} \tag{1-28}$$

这说明特征向量$\boldsymbol{x}$的坐标（$x_1^{(0)}$，$x_2^{(0)}$，…，$x_n^{(0)}$）满足线性齐次方程组

$$\begin{cases}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=\lambda_0x_1\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=\lambda_0x_2\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n=\lambda_0x_n\end{cases}$$

即

$$\begin{cases}(\lambda_0-a_{11})x_1-\quad a_{12}x_2\quad-\cdots-a_{1n}x_n=0\\ \quad -a_{21}x_1+(\lambda_0-a_{22})x_2-\cdots-a_{2n}x_n=0\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \quad -a_{n1}x_1-\cdots\cdots\cdots+(\lambda_0-a_{nn})x_n=0\end{cases} \tag{1-29}$$

由于$\boldsymbol{x}\neq 0$，所以它的坐标$x_1^{(0)}$，$x_2^{(0)}$，…，$x_n^{(0)}$不全为零，即齐次方程组（1-29）有非零解。而方程组（1-29）有非零解的充要条件是它的系数行列式为零。即

$$|\lambda_0 I_n - A| = \begin{vmatrix} \lambda_0 - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda_0 - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda_0 - a_{nn} \end{vmatrix} = 0$$

定义 1.5.2 设 A 是数域 F 上的 n 阶矩阵，λ 是一个文字，矩阵 $\lambda I_n - A$ 称为 A 的特征矩阵，行列式

$$|\lambda I_n - A| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} \qquad (1\text{-}30)$$

称为 A 的特征多项式，A 的特征多项式的根称为 A 的特征根（或特征值），以 A 的特征根 λ_0 代入方程（1-29）所得到的非零解，称为 A 的对应于 λ_0 的特征向量，简称 A 的特征向量。

矩阵 A 的特征多项式在复数范围内有 n 个根，因此在复数域内任何一个 n 阶矩阵 A 有 n 个特征值（重根包括重数）。矩阵 A 的所有相异特征值的全体称为 A 的谱，并用 $\sigma(A)$ 表示。

定理 1.5.1 相似矩阵有相同的特征多项式。

〔证明〕 设

$$B = P^{-1}AP$$

则

$$\begin{aligned} |\lambda I_n - B| &= |\lambda I_n - P^{-1}AP| = |P^{-1}(\lambda I_n - A)P| \\ &= |P^{-1}||\lambda I_n - A||P| = |\lambda I_n - A| \end{aligned}$$

】

推论 1.5.1 相似矩阵有相同的谱。

由定理 1.5.1 即可证得。

推论 1.5.2 设 $\boldsymbol{x}$ 是矩阵 A 的特征值 λ 所对应的特征向量，则 $P^{-1}\boldsymbol{x}$ 是矩阵 $B = P^{-1}AP$ 的特征值 λ 所对应的特征向量。

〔证明〕 因为

$$\begin{aligned} B(P^{-1}\boldsymbol{x}) &= P^{-1}AP(P^{-1}\boldsymbol{x}) = P^{-1}A\boldsymbol{x} \\ &= \lambda(P^{-1}\boldsymbol{x}) \end{aligned}$$

故得。 】

由定理 1.5.1 可见，线性变换的矩阵的特征多项式与基的选择无关，它是直接被线性变换决定的。因此，以后就可以把矩阵的特征多项式和特征值说成是线性变换的特征多项式和特征值。由公式（1-11）和推论 1.5.2 可见，线性变换的矩阵特征值 λ 所对应的特征向量 $\boldsymbol{x}$ 也

可以说成是线性变换的特征向量。因此，以后可以仅研究矩阵的特征值和特征向量。

定理1.5.2 设λ_1，λ_2，…，λ_r为A的相异特征值，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$是分别属于λ_1，λ_2，…，λ_r的特征向量，则$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$线性无关。

〔证明〕 设

$$c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2+\cdots+c_r\boldsymbol{x}_r=\boldsymbol{0}$$

以$\lambda_1 I_n-A$左乘上式两端，有

$$(\lambda_1-\lambda_2)c_2\boldsymbol{x}_2+(\lambda_1-\lambda_3)c_3\boldsymbol{x}_3+\cdots+(\lambda_1-\lambda_r)c_r\boldsymbol{x}_r=\boldsymbol{0}$$

再以$\lambda_2 I_n-A$左乘上式两端，有

$$(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)c_3\boldsymbol{x}_3+\cdots+(\lambda_1-\lambda_r)(\lambda_2-\lambda_r)c_r\boldsymbol{x}_r=\boldsymbol{0}$$

继续上面的过程，最后得到

$$(\lambda_1-\lambda_r)(\lambda_2-\lambda_r)\cdots(\lambda_{r-1}-\lambda_r)c_r\boldsymbol{x}_r=\boldsymbol{0}$$

因为$\lambda_i\neq\lambda_j$（$i\neq j$），$\boldsymbol{x}_r\neq\boldsymbol{0}$，所以有

$$c_r=0$$

同理可证

$$c_1=c_2=\cdots=c_{r-1}=0$$

即$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_r$线性无关。 ▌

定义1.5.3 设λ_1，λ_2，…，λ_r是A的谱，其重数分别为p_1，p_2，…，p_r，则称p_i为λ_i的**代数重复度**，齐次方程组$A\boldsymbol{x}=\lambda_i\boldsymbol{x}$的解空间称为$A$的**特征子空间**，记作$V_{\lambda_i}$。

显然，特征子空间V_{λ_i}的维数就是属于λ_i的线性无关的特征向量的最大个数，称为λ_i的**几何重复度**。

定理1.5.3 矩阵A的任一特征值的几何重复度不大于它的代数重复度。

〔证明〕 设λ_1为A的一个特征值，齐次方程组$A\boldsymbol{x}=\lambda_1\boldsymbol{x}$的基础解系为$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_k$，取$\boldsymbol{x}_{k+1}$，…，$\boldsymbol{x}_n$使得

$$\boldsymbol{x}_1,\ \cdots,\ \boldsymbol{x}_k,\ \boldsymbol{x}_{k+1},\ \cdots,\ \boldsymbol{x}_n$$

构成V的一组基，命

$$P=[\boldsymbol{x}_1,\ \cdots,\ \boldsymbol{x}_k,\ \boldsymbol{x}_{k+1},\ \cdots,\ \boldsymbol{x}_n]$$

故 $$P^{-1}P=[P^{-1}\boldsymbol{x}_1,\ P^{-1}\boldsymbol{x}_2,\ \cdots,\ P^{-1}\boldsymbol{x}_n]=I_n$$

此即 $$P^{-1}\boldsymbol{x}_1=(1,\ 0,\ \cdots,\ 0)^T,\ P^{-1}\boldsymbol{x}_2=(0,\ 1,\ 0,\ \cdots,\ 0)^T,\ \cdots$$

$$P^{-1}\boldsymbol{x}_k=(0,\ \cdots,\ 0,\ 1,\ \cdots,\ 0)^T,\ \cdots,\ P^{-1}\boldsymbol{x}_n=(0,\ \cdots,\ 0,\ 1)^T$$

于是

$$P^{-1}AP=(P^{-1}A\boldsymbol{x}_1,\ P^{-1}A\boldsymbol{x}_2,\ \cdots,\ P^{-1}A\boldsymbol{x}_k,\ \cdots,\ P^{-1}A\boldsymbol{x}_n)$$

$$=(\lambda_1 P^{-1}\boldsymbol{x}_1,\ \lambda_1 P^{-1}\boldsymbol{x}_2,\ \cdots,$$

$$\lambda_1 P^{-1}\boldsymbol{x}_k,\ P^{-1}A\boldsymbol{x}_{k+1},\ \cdots,\ P^{-1}A\boldsymbol{x}_n)$$

$$=\left(\begin{array}{ccc:c} \lambda_1 & & & \\ & \ddots & & * \\ & & \lambda_1 & \\ \hdashline & 0 & & A_1 \end{array}\right)$$

式中 0 表示 $(n-k)\times k$ 零矩阵，$*$ 表示 $k\times(n-k)$ 矩阵，A_1 表示 $n-k$ 阶矩阵。

因此

$$|\lambda I_n-A|=|\lambda I_n-P^{-1}AP|$$

$$=(\lambda-\lambda_1)^k|\lambda I_{n-k}-A_1|$$

这说明 λ_1 的几何重复度（它所对应的特征向量个数）不大于 λ_1 的代数重复度（特征多项式 $|\lambda I_n-A|$ 中因子 $\lambda-\lambda_1$ 的次数）。]

推论 1.5.3 如果矩阵 A 的特征值 λ_0 的代数重复度是 1，则 $\dim V_{\lambda_0}=1$。

证略。

定理 1.5.4 设 A 的谱为 $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_r$；$\boldsymbol{x}_{i1},\ \boldsymbol{x}_{i2},\ \cdots,\ \boldsymbol{x}_{im_i}$ 是 A 的属于特征值 λ_i 的线性无关的特征向量，则 $\boldsymbol{x}_{11},\ \boldsymbol{x}_{12},\ \cdots,\ \boldsymbol{x}_{1m_1},\ \boldsymbol{x}_{21},\ \boldsymbol{x}_{22},\ \cdots,\ \boldsymbol{x}_{2m_2},\ \cdots,\ \boldsymbol{x}_{r1},\ \boldsymbol{x}_{r2},\ \cdots,\ \boldsymbol{x}_{rm_r}$ 线性无关。

〔证明〕 设

$$c_{11}\boldsymbol{x}_{11}+\cdots+c_{1r_1}\boldsymbol{x}_{1r_1}+c_{21}\boldsymbol{x}_{21}+\cdots$$

$$+c_{r1}\boldsymbol{x}_{r_1}+\cdots+c_{rm_r}\boldsymbol{x}_{rm_r}=0$$

成立，其中 $c_{ij}\in F$（$i=1,\ 2,\ \cdots,\ r$；$j=1,\ 2,\ \cdots,\ m_i$）。对上式左乘以 $\lambda_1 I-A$，利用 $(\lambda_1 I-A)\boldsymbol{x}_{1j}=0$（$j=1,\ 2,\ \cdots,\ m_1$）与 $A\boldsymbol{x}_{ij}=\lambda_i\boldsymbol{x}_{ij}$（$i=2,3,\cdots,\ r$；$j=1,2,\cdots,\ m_r$），则

$$(\lambda_1-\lambda_2)c_{21}\boldsymbol{x}_{21}+\cdots+(\lambda_1-\lambda_r)c_{rm_r}\boldsymbol{x}_{rm_r}=0$$

依次左乘 $(\lambda_2 I-A)$，$(\lambda_3 I-A)$，$\cdots$，$(\lambda_{r-1}I-A)$，最后得

$$(\lambda_1-\lambda_r)(\lambda_2-\lambda_r)\cdots(\lambda_{r-1}-\lambda_r)(c_{r1}\boldsymbol{x}_{r1}+\cdots+c_{rm_r}\boldsymbol{x}_{rm_r})=0$$

因为

$$\lambda_j\neq\lambda_r\quad(j=1,\ 2,\ \cdots,\ r-1)$$

故得

$$c_{r1}\boldsymbol{x}_{r1}+c_{r2}\boldsymbol{x}_{r2}+\cdots+c_{rm_r}\boldsymbol{x}_{rm_r}=0$$

因为 $\boldsymbol{x}_{r1}$，$\boldsymbol{x}_{r2}$，…，$\boldsymbol{x}_{rm_r}$ 是线性独立的，所以

$$c_{r1}=c_{r2}=\cdots=c_{rm_r}=0$$

重复同样步骤，最后可以证明所有的 $c_{ij}=0$。 ▮

2. 化矩阵为对角矩阵

可以认为对角矩阵是矩阵中最简单的一种，现在我们要研究矩阵 A 与对角矩阵相似的条件。

定理 1.5.5 n 阶矩阵 A 与对角矩阵相似的充要条件是 A 有 n 个线性无关的特征向量。

〔证明〕 *必要性* 设满秩矩阵 P，满足

$$P^{-1}AP=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \tag{1-31}$$

其中 $\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 表示对角线元素分别是 λ_1，λ_2，…，λ_n 的 n 阶对角矩阵。把 P 按列向量进行分块

$$P=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) \tag{1-32}$$

将式(1-32)代入式(1-31)得

$$A(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)=(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

于是

$$A\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i \quad (i=1, 2, \cdots, n) \tag{1-33}$$

因为 P 是满秩的，所以 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_n$ 是线性无关的。从而由式(1-33)知，A 有 n 个线性无关的特征向量。

充分性 设 A 有 n 个线性无关的特征向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$，即 $A\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i (i=1,2,\cdots,n)$。命

$$P=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$$

显然 P 是满秩的。故

$$\begin{aligned}AP&=A(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)\\&=(A\boldsymbol{x}_1, A\boldsymbol{x}_2, \cdots, A\boldsymbol{x}_n)\\&=(\lambda_1\boldsymbol{x}_1, \lambda_2\boldsymbol{x}_2, \cdots, \lambda_n\boldsymbol{x}_n)\\&=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)\\&=P\,\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)\end{aligned}$$

即

$$P^{-1}AP=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$ ▮

推论 1.5.4 设 $P^{-1}AP=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$，则 λ_1，λ_2，…，λ_n 是 A 的 n 个特征值，P 的第 i 个列向量是 A 的属于特征值 λ_i 的特征向量。

证略。

定理 1.5.6 A 与对角形相似的充要条件是 A 的每一个特征值的代数重复度等于几何重复度。

〔证明〕 设 n 阶矩阵的谱为 λ_1，λ_2，…，λ_r。λ_i 的代数重复度为 p_i，几何重复度为 $m_i(i=1, 2, \cdots, r)$。

则

$$p_1+p_2+\cdots+p_r=n$$

由定理 1.5.3 知

$$m_1+m_2+\cdots+m_r\leqslant p_1+p_2+\cdots+p_r=n$$

由定理 1.5.5 知

$$m_1+m_2+\cdots+m_r=n$$

故得

$$m_1=p_1, \quad m_2=p_2, \quad \cdots, \quad m_r=p_r$$

】

我们经常称代数重复度和几何重复度相等的矩阵是**单纯矩阵**。因此定理 1.5.6 也可以叙述为：矩阵 A 与对角形相似的充要条件是 A 为单纯矩阵。

推论 1.5.5 若矩阵 A 的特征根全是单根，则 A 必与对角矩阵相似。

定理 1.5.7 设 n 阶矩阵 A 的谱为 λ_1，λ_2，…，λ_r，特征值 λ_i 的代数重复度为 p_i（$i=1, 2, \cdots, r$），则 A 与对角矩阵相似的充要条件是特征矩阵 $\lambda_i I-A$ 的秩是 $n-p_i$（$i=1, 2, \cdots, r$）。

〔证明〕 由定理 1.5.6 知 λ_i 的代数重复度 p_i 等于它的几何重复度，而 λ_i 的几何重复度就是线性齐次方程组 $(\lambda_i I-A)x=0$ 的基础解系向量个数，即 λ_i 的几何重复度等于 $n-$秩$(\lambda_i I-A)$，所以

$$p_i=n-\text{秩}(\lambda_i I-A)$$

于是

$$\text{秩}(\lambda_i I-A)=n-p_i$$

】

习　题

1-1　设 s 个 n 维向量

$$x_j=(a_{j1}, a_{j2}, \cdots, a_{jn}) \qquad j=1, 2, \cdots, s$$

适合条件

$$2|a_{jj}|>\sum_{k=1}^{n}|a_{jk}| \qquad j=1, 2, \cdots, s$$

试证：x_1，x_2，…，x_s 线性无关。

1-2　试证：所有 n 阶对称矩阵构成 $n(n+1)/2$ 维线性空间；所有 n 阶反对称矩阵构成 $n(n-1)/2$ 维线性空间。

1-3　在 R^4 中，求向量 $\boldsymbol{x}=(1,\ 2,\ 1,\ 1)$ 在基

$$\boldsymbol{x}_1=(1,\ 1,\ 1,\ 1),\qquad \boldsymbol{x}_2=(1,1,-1,-1)$$

$$\boldsymbol{x}_3=(1,\ -1,\ 1,\ -1),\ \boldsymbol{x}_4=(1,\ -1,\ -1,\ 1)$$

下的坐标。

1-4　在 R^4 中求由基 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，$\boldsymbol{x}_3$，$\boldsymbol{x}_4$ 到基 $\boldsymbol{y}_1$，$\boldsymbol{y}_2$，$\boldsymbol{y}_3$，$\boldsymbol{y}_4$ 的过渡矩阵，并求向量 $\boldsymbol{x}$ 在所指定基下的坐标。设

$$\boldsymbol{x}_1=(1,\ 0,\ 0,\ 0),\ \boldsymbol{x}_2=(0,\ 1,\ 0,\ 0)$$

$$\boldsymbol{x}_3=(0,\ 0,\ 1,\ 0),\ \boldsymbol{x}_4=(0,\ 0,\ 0,\ 1)$$

与

$$\boldsymbol{y}_1=(2,\ 1,\ -1,\ 1),\qquad \boldsymbol{y}_2=(0,\ 3,\ 1,\ 0)$$

$$\boldsymbol{y}_3=(5,\ 3,\ 2,\ 1),\qquad \boldsymbol{y}_4=(6,\ 6,\ 1,\ 3)$$

$\boldsymbol{x}=(x_1,\ x_2,\ x_3,\ x_4)$ 在基 $\boldsymbol{y}_1$，$\boldsymbol{y}_2$，$\boldsymbol{y}_3$，$\boldsymbol{y}_4$ 下的坐标。

1-5　设

$$A=\begin{pmatrix}1&0&0\\0&1&0\\3&1&2\end{pmatrix}$$

求三阶矩阵中全体与 A 可交换的矩阵所成的子空间的维数和一组基。

1-6　求由向量 $\boldsymbol{x}_i$ 生成的子空间与由向量 $\boldsymbol{y}_i$ 生成的子空间的交的基和维数。设

（1）　$\begin{cases}\boldsymbol{x}_1=(1,\ 2,\ 1,\ 0)\\ \boldsymbol{x}_2=(-1,\ 1,\ 1,\ 1)\end{cases}\qquad \begin{cases}\boldsymbol{y}_1=(2,\ -1,\ 0,\ 1)\\ \boldsymbol{y}_2=(1,\ -1,\ 3,\ 7)\end{cases}$

（2）　$\begin{cases}\boldsymbol{x}_1=(1,\ 2,\ -1,\ -2)\\ \boldsymbol{x}_2=(3,\ 1,\ 1,\ 1)\\ \boldsymbol{x}_3=(-1,\ 0,\ 1,\ -1)\end{cases}\qquad \begin{cases}\boldsymbol{y}_1=(2,\ 5,\ -6,\ -5)\\ \boldsymbol{y}_2=(-1,\ 2,\ -7,\ 3)\end{cases}$

1-7　判别下面所定义的变换，哪些是线性的，哪些不是：

（1）　在 R^3 中：$\mathscr{A}(x_1,\ x_2,\ x_3)=(x_1^2,\ x_2+x_3,\ x_3^2)$

（2）　在 R^3 中：$\mathscr{A}(x_1,\ x_2,\ x_3)=(2x_1-x_2,\ x_2+x_3,\ x_1)$

（3）　在 $F(x)$ 中：$\mathscr{A}f(x)=f(x+1)$

1-8　设 $\mathscr{A}$ 是 n 维线性空间 V 中的一个线性变换，对某个 $\boldsymbol{x}\in V$，$\mathscr{A}^{k-1}(\boldsymbol{x})\neq 0$，$\mathscr{A}^k(\boldsymbol{x})=0$。试证：$\boldsymbol{x}$，$\mathscr{A}(\boldsymbol{x})$，$\mathscr{A}^2(\boldsymbol{x})$，…，$\mathscr{A}^{k-1}(\boldsymbol{x})$ 线性无关（$k>0$）。

1-9　若 n 维线性空间 V 中线性变换 $\mathscr{A}$ 使对 V 中任何向量 $\boldsymbol{x}$，有 $\mathscr{A}^{n-1}(\boldsymbol{x})\neq 0$，但

$\mathscr{A}^n(x)=0$，求$\mathscr{A}$在某一组基下的矩阵表示。

1-10 设$\mathscr{A}$是线性空间V的线性变换，它在V中基x_1，x_2，x_3下的矩阵表示是

$$A=\begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & 3 \\ 2 & 1 & 5 \end{pmatrix}$$

求$\mathscr{A}$在基$y_1=x_1$，$y_2=x_1+x_2$，$y_3=x_1+x_2+x_3$下的矩阵表示。

1-11 设$\mathscr{A}$在基$x_1=(-1, 1, 1)$，$x_2=(1, 0, -1)$，$x_3=(0, 1, 1)$下的矩阵表示是

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix}$$

求$\mathscr{A}$在基$e_1=(1, 0, 0)$，$e_2=(0, 1, 0)$，$e_3=(0, 0, 1)$下的矩阵表示。

1-12 设x_1，x_2，x_3，x_4是四维线性空间V的一组基，已知线性变换$\mathscr{A}$在这组基下的矩阵表示为

$$\begin{pmatrix} 1 & 0 & 2 & 1 \\ -1 & 2 & 1 & 3 \\ 1 & 2 & 5 & 5 \\ 2 & -2 & 1 & -2 \end{pmatrix}$$

（1） 求$\mathscr{A}$在基$y_1=x_1-2x_2+x_4$，$y_2=3x_2-x_3-x_4$，$y_3=x_3+x_4$，$y_4=2x_4$下的矩阵表示。

（2） 求$\mathscr{A}$的核与值域。

（3） 在$\mathscr{A}$的核中选一组基，把它扩充成V的一组基，并求$\mathscr{A}$在这组基下的矩阵表示。

（4） 在$\mathscr{A}$的值域中选一组基，把它扩充成V的一组基，并求$\mathscr{A}$在这组基下的矩阵表示。

1-13 求三维空间关于用矩阵

$$\begin{pmatrix} 4 & -2 & 2 \\ -2 & 0 & 2 \\ -1 & 1 & 1 \end{pmatrix}$$

所给出的线性变换不变的所有子空间。

1-14 求下列矩阵的特征值和特征向量。

（1）
$$\begin{pmatrix} 2 & -1 & 2 \\ 5 & -3 & 3 \\ -1 & 0 & 2 \end{pmatrix}$$
（2）
$$\begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix}$$
（3）
$$\begin{pmatrix} 4 & -5 & 2 \\ 5 & -7 & 3 \\ 6 & -9 & 4 \end{pmatrix}$$
（4）
$$\begin{pmatrix} 4 & -5 & 7 \\ 1 & -4 & 9 \\ -4 & 0 & 5 \end{pmatrix}$$
（5）
$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$
（6）
$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

1-15　设$\mathscr{A}$是n阶可逆矩阵，则AB与BA相似。

第二章　λ-矩阵与标准形

这一章讨论λ-矩阵和数字矩阵的相似标准形。内容包括λ-矩阵的概念，λ-矩阵的标准形，λ-矩阵的除法，数字矩阵相似的条件，矩阵的有理标准形和Jordan标准形。

§2.1　λ-矩阵的概念

定义2.1.1　设$a_{ij}(\lambda)(i=1, 2, \cdots, m; j=1, 2, \cdots, n)$为数域$F$上的多项式，则称以$a_{ij}(\lambda)$为元素的$m\times n$矩阵

$$A(\lambda)=\begin{pmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ a_{21}(\lambda) & a_{22}(\lambda) & \cdots & a_{2n}(\lambda) \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots & a_{mn}(\lambda) \end{pmatrix}$$

为**多项式矩阵**或**λ-矩阵**。

数字矩阵和特征矩阵$\lambda I-A$均为λ-矩阵的特例。

λ-矩阵的加法、数乘和乘法运算与数字矩阵相同，而且有相同的运算规律，这些就不再重复叙述和证明。

定义 2.1.2　如果λ-矩阵$A(\lambda)$中有一个r阶（$r\geqslant 1$）子式不为零，而所有$r+1$阶子式（如果有的话）全为零，则称$A(\lambda)$的**秩**为r。零矩阵的秩为0。

定义2.1.2从字面上来看与数字矩阵相同，但是要注意到$A(\lambda)$的行列式及一切子式是λ的多项式，所以$r+1$阶子式为零的涵义是$r+1$阶子式恒为零。

定义 2.1.3　一个n阶λ-矩阵$A(\lambda)$称为**可逆的**，如果有一个n阶λ-矩阵$B(\lambda)$，使得

$$A(\lambda)B(\lambda)=B(\lambda)A(\lambda)=I \tag{2-1}$$

这里I是n阶单位矩阵。$B(\lambda)$称为$A(\lambda)$的**逆矩阵**，记为$A^{-1}(\lambda)$。

定理 2.1.1　一个n阶λ-矩阵$A(\lambda)$可逆的充要条件是$\det A(\lambda)$是一个非零的常数。

〔证明〕　必要性　设$A(\lambda)$可逆，在式（2-1）的两边取行列式

$$|A(\lambda)||B(\lambda)|=1$$

因为$|A(\lambda)|$与$|B(\lambda)|$都是λ的多项式，所以根据它们乘积是1可以推知，它们都是零次多项式，此即$|A(\lambda)|$是非零的数。

充分性　设

$$d=|A(\lambda)|$$

是一个非零的数。矩阵

$$\frac{1}{d}A^*(\lambda) \qquad (2\text{-}2)$$

是一个 λ-矩阵，其中 $A^*(\lambda)$ 是 $A(\lambda)$ 的伴随矩阵*，所以

$$A(\lambda)\frac{1}{d}A^*(\lambda)=\frac{1}{d}A(\lambda)A^*(\lambda)=I$$

因此$A(\lambda)$可逆，且它的逆矩阵是$\frac{1}{d}A^*(\lambda)$。 ▌

定义 2.1.4 下列各种类型的变换，叫做 λ-矩阵的初等变换：

（1） 矩阵的任二行（列）互换位置。

（2） 非零常数 a 乘矩阵的某一行（列）。

（3） 矩阵的某一行（列）的$\varphi(\lambda)$倍加到另一行（列）上去，其中$\varphi(\lambda)$是λ 的一个多项式。

对单位矩阵施行上述三种类型的初等变换便得相应的三种λ-矩阵的初等矩阵 $P(i,j)$，$P(i(c))$，$P(i,j(\varphi))$即

$$P(i,j)=\begin{pmatrix}1&&&&&\\&\ddots&&&&\\&&0&1&&\\&&&\ddots&&\\&&1&0&&\\&&&&\ddots&\\&&&&&1\end{pmatrix}\begin{matrix}\\\\-i\text{ 行}\\\\-j\text{ 行}\\\\\\\end{matrix}$$

$$P(i(c))=\begin{pmatrix}1&&&&\\&\ddots&&&\\&&c&&\\&&&\ddots&\\&&&&1\end{pmatrix}\begin{matrix}\\\\-i\text{ 行}\\\\\\\end{matrix}$$

$$P(i,j(\varphi))=\begin{pmatrix}1&&&\\&\ddots&\varphi(\lambda)&\\&&\ddots&\\&&&1\end{pmatrix}\begin{matrix}-i\text{ 行}\\\\-j\text{ 行}\\\\\end{matrix}$$

对一个$m\times n$ λ-矩阵$A(\lambda)$的行作一初等变换，相当于用相应的m阶初等矩阵左乘$A(\lambda)$。对$A(\lambda)$的列作一初等变换，相当于用相应的 n 阶初等矩阵右乘$A(\lambda)$。

容易验证，初等矩阵都是可逆的，并且

$$P(i,j)^{-1}=P(i,j),\qquad P(i(c))^{-1}=P(i(c^{-1}))$$

$$P(i,j(\varphi))^{-1}=P(i,j(-\varphi))$$

定义 2.1.4 如果 $A(\lambda)$ 经过有限次的初等变换后变成 $B(\lambda)$，则称 $A(\lambda)$ 与 $B(\lambda)$ 等

---

* $A^*(\lambda)$的定义与数字矩阵类似。

价，记之为$A(\lambda)\simeq B(\lambda)$。

定理 2.1.2 $A(\lambda)\simeq B(\lambda)$的充要条件是存在两个可逆矩阵 $P(\lambda)$ 与$Q(\lambda)$，使得

$$B(\lambda)=P(\lambda)A(\lambda)Q(\lambda) \tag{2-3}$$

〔证明〕 由定义 2.1.4 知 $A(\lambda)$与 $B(\lambda)$ 等价的充要条件是存在一系列的初等矩阵P_1，P_2，…，P_r与Q_1，Q_2，…Q_t，使得

$$B(\lambda)=P_rP_{r-1}\cdots P_2P_1A(\lambda)Q_1Q_2\cdots Q_t$$

若命

$$P(\lambda)=P_rP_{r-1}\cdots P_2P_1,$$

$$Q(\lambda)=Q_1Q_2\cdots Q_t$$

因为初等矩阵是可逆矩阵，所以 $P(\lambda)$，$Q(\lambda)$ 均可逆。于是

$$B(\lambda)=P(\lambda)A(\lambda)Q(\lambda)$$

λ-矩阵的等价关系满足 ▮

（1） 自反性 每一个 λ- 矩阵与自己等价。

（2） 对称性 若$A(\lambda)\simeq B(\lambda)$，则$B(\lambda)\simeq A(\lambda)$。

（3） 传递性 若$A(\lambda)\simeq B(\lambda)$，$B(\lambda)\simeq C(\lambda)$，则 $A(\lambda)\simeq C(\lambda)$。

因此可以把所有的 $m\times n$阶λ- 矩阵按等价关系分类，叫做等价类，使在同一类中的矩阵是等价的，在不同类中的矩阵是不等价的。所以我们只要研究每一类矩阵的标准形。

§2.2 λ-矩阵的标准形

1. λ-矩阵的标准形

这一节主要证明任何一个 λ- 矩阵等价于对角矩阵以及标准形是唯一的。

引理 2.2.1 设 λ-矩阵 $A(\lambda)$ 的左上角元素 $a_{11}(\lambda)\neq 0$，并且 $A(\lambda)$ 中至少有一个元素不能被它整除，那么一定可以找到一个与 $A(\lambda)$ 等价的矩阵 $B(\lambda)$，它的左上角元素也不为零，但是次数比 $a_{11}(\lambda)$ 的次数低。

〔证明〕 根据 $A(\lambda)$ 中不能被 $a_{11}(\lambda)$ 整除的元素所在的位置，分三种情况讨论。

（1） 若在 $A(\lambda)$ 的第一列中有一个元素 $a_{i1}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除，即有

$$a_{i1}(\lambda)=a_{11}(\lambda)q(\lambda)+r(\lambda)$$

其中余式 $r(\lambda)\neq 0$，且次数比 $a_{11}(\lambda)$ 的次数低。

对 $A(\lambda)$ 作两次初等行变换，首先第一行乘以 $-q(\lambda)$ 加到第 i 行，第 i 行第一列的元素成为 $r(\lambda)$，然后把第一行和第 i 行互换得到新的 λ-矩阵 $B(\lambda)$，$B(\lambda)$ 左上角元素为 $r(\lambda)$，故为引理所求矩阵。

（2） 在 $A(\lambda)$ 的第一行中有一个元素$a_{1i}(\lambda)$不能被 $a_{11}(\lambda)$ 整除，这种情况的证明与情况 1 类似。

（3） $A(\lambda)$的第一行与第一列中的元素都可以被 $a_{11}(\lambda)$ 整除，但 $A(\lambda)$ 中有一个元

素$a_{ij}(\lambda)$（$i>1,j>1$）不能被$a_{11}(\lambda)$整除。我们设$a_{i1}(\lambda)=a_{11}(\lambda)\varphi(\lambda)$，对$A(\lambda)$作两次初等行变换，首先第一行乘以$-\varphi(\lambda)$加到第$i$行，第$i$行第一列的元素变为0，第$i$行第$j$列的元素变为$a_{ij}(\lambda)-a_{1j}(\lambda)\varphi(\lambda)$；其次把第$i$行的元素乘以$-1$加到第一行，第一行第一列的元素仍为$a_{11}(\lambda)$，第一行第$j$列的元素变为$a_{ij}(\lambda)+[1-\varphi(\lambda)]a_{1j}(\lambda)$，它不能被$a_{11}(\lambda)$所整除，这就化为已经证明了的情况2。 ∎

定理 2.2.1 任意一个非零的$m\times n$ λ-矩阵$A(\lambda)$都等价于一个对角矩阵，即

$$A(\lambda)\simeq\begin{pmatrix} d_1(\lambda) & & & & & \\ & d_2(\lambda) & & & & \\ & & \ddots & & & \\ & & & d_r(\lambda) & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix} \qquad (2\text{-}4)$$

其中$r\geqslant1$，$d_i(\lambda)$是首项系数为1的多项式，且

$$d_i(\lambda)|d_{i+1}(\lambda) \qquad (i=1,\ 2,\ \cdots,\ r-1)$$

记号$d_i(\lambda)|d_{i+1}(\lambda)$表示$d_{i+1}(\lambda)$能被$d_i(\lambda)$整除。

〔证明〕 设$a_{11}(\lambda)\neq0$，否则，总可以经过行列调整，使得$A(\lambda)$的左上角元素不为零。若$a_{11}(\lambda)$不能整除$A(\lambda)$的所有元素，由引理2.2.1，可以找到与$A(\lambda)$等价的$B_1(\lambda)$，它的左上角元素$b_1(\lambda)\neq0$，并且次数比$a_{11}(\lambda)$低。如果$b_1(\lambda)$还不能整除$B_1(\lambda)$的所有元素，由引理2.2.1又可以找到与$B_1(\lambda)$等价的$B_2(\lambda)$，它的左上角元素$b_2(\lambda)\neq0$，并且次数比$b_1(\lambda)$低，如果$b_2(\lambda)$还不能整除$B_2(\lambda)$的所有元素，继续上述步骤，得到一系列彼此等价的λ-矩阵$A(\lambda)$，$B_1(\lambda)$，…。它们的左上角元素皆不为零，而且次数越来越低。但是多项式的次数是非负整数，不可能无止境地降低，因此在有限步之后，就会得到一个λ-矩阵$B(\lambda)$，它的左上角元素$b_s(\lambda)\neq0$，而且可以整除$B_s(\lambda)$的全部元素$b_{ij}(\lambda)$，即

$$b_{ij}(\lambda)=b_s(\lambda)q_{ij}(\lambda)$$

显然，可对$B_s(\lambda)$分别作一系列的初等行变换与初等列变换，使得第一行除左上角元素$b_s(\lambda)$外全为零，第一列除左上角元素$b_s(\lambda)$外全为零。即

$$B_s(\lambda)\simeq\begin{pmatrix} b_s(\lambda) & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & A_1(\lambda) & \\ 0 & & & \end{pmatrix}$$

因为$A_1(\lambda)$的元素是$B_s(\lambda)$元素的组合，而$B_s(\lambda)$的$b_s(\lambda)$可以整除$B_s(\lambda)$的所有元素，所以$b_s(\lambda)$可以整除$A_1(\lambda)$的所有元素。如果$A_1(\lambda)\neq0$，则对于$A_1(\lambda)$可以重复上述过程，进而把矩阵化成

$$\begin{pmatrix} d_1(\lambda) & 0 & \cdots & & 0 \\ 0 & d_2(\lambda) & 0 & \cdots & 0 \\ \vdots & 0 & & & \\ 0 & \vdots & A_2(\lambda) & & \\ 0 & 0 & & & \end{pmatrix}$$

其中 $d_1(\lambda)$ 与 $d_2(\lambda)$ 都是首项系数为 1 的多项式（$d_1(\lambda)$与$b_s(\lambda)$只差一个常数倍数)，而且 $d_1(\lambda)|d_2(\lambda)$，$d_2(\lambda)$能整除$A_2(\lambda)$的所有元素。如此下去，$A(\lambda)$最后化成了所要求的形式。】

2. 不变因子与初等因子

在化 λ-矩阵为对角矩阵的问题中，不变因子和初等因子起着重要作用。

定义 2.2.1 $A(\lambda)$ 最后化成的式（2-4）的右端，称为$A(\lambda)$的**标准形**。$d_1(\lambda)$，$d_2(\lambda)$，…，$d\ (\lambda)$ 称为 λ-矩阵$A(\lambda)$的**不变因子**。

为了研究 λ-矩阵标准形的唯一性，还需研究行列式因子。

定义 2.2.2 设 λ-矩阵的秩为 r，对于正整数 k，$1\leqslant k\leqslant r$，$A(\lambda)$中必有非零的 k 级子式，$A(\lambda)$ 中全部 k 级子式的最高公因式 $D_k(\lambda)$ 称为 $A(\lambda)$ 的 k **级行列式因子**。

由定义 2.2.2 可知，对于秩为 r 的 λ-矩阵，行列式因子一共有 r 个。关于行列式因子有下面重要结论。

定理 2.2.2 等价矩阵具有相同的秩与相同的各级行列式因子。

〔证明〕 我们只要证明，λ-矩阵经过一次初等变换，秩与行列式因子是不变的。

设 λ-矩阵 $A(\lambda)$ 经过一次初等行变换变成 $B(\lambda)$，$D_k(\lambda)$ 与 $\bar{D}_k(\lambda)$ 分别是 $A(\lambda)$ 与 $B(\lambda)$ 的 k 级行列式因子。我们来证明 $D_k(\lambda)=\bar{D}_k(\lambda)$，分三种情况讨论。

（1） $B(\lambda)$是由 $A(\lambda)$交换某两行而得到。这时 $B(\lambda)$ 的每个 k 级子式或者等于 $A(\lambda)$ 的某个 k 级子式，或者与 $A(\lambda)$ 的某一个 k 级子式相差一个符号，因此 $D_k(\lambda)$ 是 $B(\lambda)$ 的 k 级子式的公因式，从而$D_k(\lambda)|\bar{D}_k(\lambda)$。

（2） $B(\lambda)$ 是由 $A(\lambda)$ 的某一行乘以非零数 C 而得。这时 $B(\lambda)$ 的每个 k 级子式或者等于 $A(\lambda)$ 的某个 k 级子式，或者等于 $A(\lambda)$ 的某一个 k 级子式的 C 倍。因此$D_k(\lambda)$是 $B(\lambda)$的 k 级子式的公因式。从而 $D_k(\lambda)|\bar{D}_k(\lambda)$。

（3） $B(\lambda)$是由$A(\lambda)$第 j 行的$\varphi(\lambda)$倍加到第 i 行上得到，其中 $\varphi(\lambda)$ 为 λ的某个多项式。这时$B(\lambda)$ 中那些包含 i 行与 j 行的 k 级子式和那些不包含 i 行的 k 级子式都等于 $A(\lambda)$中对应的k级子式；$B(\lambda)$ 中那些包含 i 行但不包含 j 行的 k 级子式，按 i 行分成两部分之和，一部分恰等于$A(\lambda)$的一个 k 级子式，另一部分是$A(\lambda)$的另一个 k 级子式的$\pm\varphi(\lambda)$倍，也就是 $A(\lambda)$ 的两个 k 级子式的组合。因此 $D_k(\lambda)$ 是 $B(\lambda)$ 的 k 级子式的公因式，从而 $D_k(\lambda)|\bar{D}_k(\lambda)$。

对于列变换，可以完全一样地讨论。总之，如果 $A(\lambda)$ 经过一次初等变换变成$B(\lambda)$，那么$D_k(\lambda)|\bar{D}_k(\lambda)$。又由初等变换的可逆性，$B(\lambda)$ 也可以经过一次初等变换变成$A(\lambda)$。由上面的讨论，同样应有 $\bar{D}_k(\lambda)|D_k(\lambda)$，于是$D_k(\lambda)=\bar{D}_k(\lambda)$。

当 $A(\lambda)$ 的全部 k 级子式为零时，$D_k(\lambda)=0$，于是 $\bar{D}_k(\lambda)=0$，$B(\lambda)$的全部 k 级子式也

就等于零；反之亦然。因此 $A(\lambda)$ 与 $B(\lambda)$ 既有相同的各级行列式因子，又有相同的秩。]

设 λ-矩阵的标准形为

$$\begin{pmatrix} d_1(\lambda) & & & & & \\ & d_2(\lambda) & & & & \\ & & \ddots & & & \\ & & & d_r(\lambda) & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix}$$

其中 $d_1(\lambda)$, $d_2(\lambda)$, …, $d_r(\lambda)$ 是首项系数为 1 的多项式，且 $d_i(\lambda)|d_{i+1}(\lambda)$ $(i=1, 2, \cdots, r-1)$。不难知道，k 级行列式因子为

$$D_k(\lambda)=d_1(\lambda)d_2(\lambda)\cdots d_k(\lambda) \tag{2-5}$$

定理 2.2.3 λ-矩阵的标准形是唯一的。

〔证明〕 由式（2-5）知，$A(\lambda)$ 的不变因子为

$$d_1(\lambda)=D_1(\lambda),\quad d_2(\lambda)=\frac{D_2(\lambda)}{D_1(\lambda)},\ \cdots,$$

$$d_r(\lambda)=\frac{D_r(\lambda)}{D_{r-1}(\lambda)} \tag{2-6}$$

这说明 $A(\lambda)$ 的不变因子由 $A(\lambda)$ 的行列式因子唯一确定，因此 $A(\lambda)$ 的标准形是唯一的。]

利用 λ-矩阵的标准形，容易证明下述几个定理。

定理2.2.4 λ-矩阵 $A(\lambda)$ 与 $B(\lambda)$ 等价的充要条件是对于任何的 k，它们的 k 阶行列式因子相同。

〔证明〕 由定理 2.2.2 知，必要性成立。现证充分性。首先将 $A(\lambda)$ 与 $B(\lambda)$ 都化成标准形，得到它们的不变因子 $d_i(\lambda)$ 与 $\bar{d}_i(\lambda)$。于是由式（2-5）知

$$D_i(\lambda)=d_1(\lambda)d_2(\lambda)\cdots d_i(\lambda)$$

$$\bar{D}_i(\lambda)=\bar{d}_1(\lambda)\bar{d}_2(\lambda)\cdots\bar{d}_i(\lambda)$$

根据假设，$D_i(\lambda)=\bar{D}_i(\lambda)$，$(i=1, 2, \cdots)$，因此 $A(\lambda)$ 与 $B(\lambda)$ 之秩相等，否则，若 $r_A<r_B$（r_A, r_B 表示 A 与 B 之秩），则 $\bar{D}_{r_B}(\lambda)$ 有意义，而 $D_{r_A}(\lambda)$ 没有意义，矛盾！设 $A(\lambda)$ 与 $B(\lambda)$ 的秩为 r，于是它们有 r 个不变因子，则由 $D_i(\lambda)=\bar{D}_i(\lambda)$ $(i=1, 2, \cdots, r)$ 与式（2-5）便得 $d_i(\lambda)=\bar{d}_i(\lambda)$ $(i=1, 2, \cdots, r)$，即 $A(\lambda)$ 与 $B(\lambda)$ 都和同一个标准形等价，因此 $A(\lambda)\simeq B(\lambda)$。]

定理 2.2.5 $A(\lambda)\simeq B(\lambda)$ 的充要条件是 $A(\lambda)$ 与 $B(\lambda)$ 有相同的不变因子。

定理的证明留给读者作为练习。

定理 2.2.5 表明，在求 λ-矩阵的标准形时，可以利用初等变换的方法，而避免计算行

列式。

定理 2.2.4 和定理 2.2.5 都已隐含秩相等的条件。特别地，当 n 阶 λ-矩阵 $A(\lambda)$ 为满秩时，由初等变换的定义知，$\det A(\lambda)=cd_1(\lambda)\cdots d_n(\lambda)$，其中 c 为不等于零的一个常数。这表明每个不变因子 $d_i(\lambda)$ 是行列式 $\det A(\lambda)$ 的因子，又不变因子 $d_i(\lambda)$ 是由矩阵 $A(\lambda)$ 唯一确定，故它们是 $A(\lambda)$ 的不变量，这也正是称 $d_i(\lambda)$ 为不变因子的由来。

推论 2.2.1 λ-矩阵 $A(\lambda)$ 可逆的充要条件是 $A(\lambda)$ 与单位矩阵等价。

〔证明〕 *必要性* 设 $A(\lambda)$ 为一个 n 阶可逆矩阵，则由定理 2.1.1 知

$$|A(\lambda)|=d\neq 0$$

即 $A(\lambda)$ 的 n 阶行列式因子

$$D_n(\lambda)=1$$

由式（2-5）知，有关系

$$D_k(\lambda)|D_{k+1}(\lambda) \quad (k=1,\ 2,\ \cdots,\ n-1)$$

故得

$$D_k(\lambda)=1 \quad (k=1,\ 2,\ \cdots,\ n)$$

于是

$$d_k(\lambda)=1 \quad (k=1,\ 2,\ \cdots,n)$$

这说明 $A(\lambda)$ 的标准形为单位矩阵。

充分性 设

$$A(\lambda)\simeq I_n$$

所以 $A(\lambda)$ 的行列式是一个非零的常数，由定理 2.2.1 知 $A(\lambda)$ 可逆。 ▮

推论 2.2.2 λ-矩阵 $A(\lambda)$ 可逆的充要条件是 $A(\lambda)$ 可以表成一系列初等矩阵的乘积。

〔证明〕 由推论 2.2.1 知 $A(\lambda)$ 可逆的充要条件是

$$A(\lambda)\simeq I_n$$

而 $A(\lambda)$ 与 I_n 等价的充要条件是有一系列初等矩阵 P_1，$P_2,\cdots,P_l$，Q_1，Q_2，$\cdots$，Q_t 使得

$$A(\lambda)=P_1P_2\cdots P_lI_nQ_1\cdots Q_t=P_1\cdots P_lQ_1\cdots Q_t$$ ▮

由前面叙述可知，λ-矩阵的行列式因子 $D_i(\lambda)$ 与不变因子 $d_k(\lambda)$ 都是 λ 的多项式，它们都是由 $A(\lambda)$ 的元素用初等变换的方法得到。因此运算法则可限制为"加、减、乘、除"的四则运算。于是 $D_i(\lambda)$ 与 $d_i(\lambda)$ 的系数均限制在 $A(\lambda)$ 元素系数的域内。如果允许利用复数，那么每个不变因子 $d_i(\lambda)$ 可以分解为一次因式的幂积，令

$$d_1(\lambda)=(\lambda-\lambda_1)^{k_{11}}(\lambda-\lambda_2)^{k_{12}}\cdots(\lambda-\lambda_t)^{k_{1t}}$$

$$d_2(\lambda)=(\lambda-\lambda_1)^{k_{21}}(\lambda-\lambda_2)^{k_{22}}\cdots(\lambda-\lambda_t)^{k_{2t}}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$d_r(\lambda)=(\lambda-\lambda_1)^{k_{r1}}(\lambda-\lambda_2)^{k_{r2}}\cdots(\lambda-\lambda_t)^{k_{rt}}$$

因为

$$d_{i-1}(\lambda)|d_i(\lambda) \qquad (i=2,\ 3,\ \cdots,\ r)$$

所以

$$k_{1j}\leqslant k_{2j}\leqslant\cdots\leqslant k_{rj} \qquad (j=1,\ 2,\ \cdots,\ t)$$

这里的 λ_1，λ_2，…，λ_t 是 $d_r(\lambda)$ 的一切相异的根，所以 k_{r1}，k_{r2}，…，k_{rt} 无一为零，但 k_{1j}，k_{2j}，…，$k_{r-1,j}$ 中($j=1$，2，…，t)可能有为零的，而且若有 $k_{ij}=0$ ($j=1$，2，…，t；$i=1$，2，…，$r-1$)，那么也必有 $k_{1j}=k_{2j}=\cdots=k_{i-1,j}=0$。我们把

$$\begin{cases}(\lambda-\lambda_1)^{k_{11}},(\lambda-\lambda_2)^{k_{12}},\ \cdots,\ (\lambda-\lambda_t)^{k_{1t}}\\ (\lambda-\lambda_1)^{k_{21}},\ (\lambda-\lambda_2)^{k_{22}},\ \cdots,\ (\lambda-\lambda_t)^{k_{2t}}\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ (\lambda-\lambda_1)^{k_{r1}},\ (\lambda-\lambda_2)^{k_{r2}},\ \cdots,\ (\lambda-\lambda_t)^{k_{rt}}\end{cases} \tag{2-7}$$

中不是常数 1 的因式全体叫做 $A(\lambda)$ 的**初等因子**。

例如，λ-矩阵的不变因子为

$$1,\ 1,\ (\lambda-2)^5(\lambda-3)^3,\ (\lambda-5)^5(\lambda-3)^4(\lambda+2)$$

它的初等因子是

$$(\lambda-2)^5,\ (\lambda-3)^3,\ (\lambda-2)^5,\ (\lambda-3)^4,\ (\lambda+2)$$

若两个 λ-矩阵 $A(\lambda)$ 与 $B(\lambda)$ 等价，根据定理 2.2.5 它们有相同的不变因子，因此它们的初等因子也相同。但定理 2.2.5 的逆定理不一定成立，即两个 λ-矩阵的初等因子相同，它们可能不等价。例如

$$A(\lambda)=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & \lambda-4 & 0 & 0\\ 0 & 0 & (\lambda-4)^2 & 0\end{pmatrix}$$

$$B(\lambda)=\begin{pmatrix}\lambda-4 & 0 & 0 & 0\\ 0 & (\lambda-4)^2 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}$$

都是 3×4 λ-矩阵，它们的初等因子都是 $(\lambda-4)$，$(\lambda-4)^2$，但是它们的秩不相等，于是 $A(\lambda)$ 与 $B(\lambda)$ 不等价。

定理 2.2.6 $m\times n$ λ-矩阵 $A(\lambda)$ 与 $B(\lambda)$ 等价的充要条件是它们的秩相等，并且有相同的初等因子。

〔证明〕 必要性显然。现证充分性。设$A(\lambda)$与$B(\lambda)$的秩都为r，并都有式(2-7)的初等因子，其中$k_{1j}\leqslant k_{2j}\leqslant\cdots\leqslant k_{rj}$（$j=1,2,\cdots,t$）。由初等因子定义知，$A(\lambda)$与$B(\lambda)$的$r$阶不变因子$d_r(\lambda)$与$\bar{d}_r(\lambda)$相等，即

$$d_r(\lambda)=(\lambda-\lambda_1)^{k_{r1}}(\lambda-\lambda_2)^{k_{r2}}\cdots(\lambda-\lambda_t)^{k_{rt}}=\bar{d}_r(\lambda)$$

同样，对于任意的k（$1\leqslant k\leqslant r$）阶不变因子有

$$d_k(\lambda)=\bar{d}_k(\lambda)$$

因此，根据定理2.2.5有$A(\lambda)\simeq B(\lambda)$。 ▮

对于准对角形矩阵

$$A(\lambda)=\begin{pmatrix}B(\lambda) & 0\\ 0 & C(\lambda)\end{pmatrix}$$

不能从$B(\lambda)$与$C(\lambda)$的不变因子求得$A(\lambda)$的不变因子，但是能从$B(\lambda)$与$C(\lambda)$的初等因子立即得到$A(\lambda)$的初等因子。此即

定理2.2.7 设λ-矩阵

$$A(\lambda)=\begin{pmatrix}B(\lambda) & 0\\ 0 & C(\lambda)\end{pmatrix}$$

为准对角形矩阵，则$B(\lambda)$、$C(\lambda)$的各个初等因子之全体是$A(\lambda)$的全部初等因子。

〔证明〕 先将$B(\lambda)$和$C(\lambda)$分别化成标准形

$$B(\lambda)\simeq\begin{pmatrix}d_1(\lambda) & & & & & & \\ & d_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & d_{r_1}(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0\end{pmatrix}$$

$$C(\lambda)=\begin{pmatrix}\bar{d}_1(\lambda) & & & & & & \\ & \bar{d}_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & \bar{d}_{r_2}(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0\end{pmatrix}$$

$A(\lambda)$的秩$r=r_1+r_2$。把$d_i(\lambda)$和$\bar{d}_i(\lambda)$分解为不同的一次因式的幂积，即

$$d_i(\lambda)=(\lambda-\lambda_1)^{e_{i1}}(\lambda-\lambda_2)^{e_{i2}}\cdots(\lambda-\lambda_t)^{e_{it}}$$

$$\bar{d}_j(\lambda)=(\lambda-\lambda_1)^{h_{j1}}(\lambda-\lambda_2)^{h_{j2}}\cdots(\lambda-\lambda_t)^{h_{jt}}$$

$$(i=1,\ 2,\ r_1;\ j=1,\ 2,\ \cdots,\ r_2)$$

因此，$B(\lambda)$和$C(\lambda)$的初等因子分别是

$$(\lambda-\lambda_1)^{e_{i1}},\ (\lambda-\lambda_2)^{e_{i2}},\ \cdots,\ (\lambda-\lambda_t)^{e_{it}}$$

和

$$(\lambda-\lambda_1)^{h_{j1}},(\lambda-\lambda_2)^{h_{j2}},\cdots,(\lambda-\lambda_t)^{h_{jt}}$$

中不为常数的多项式。

现在证明$B(\lambda)$、$C(\lambda)$的这些初等因子就是$A(\lambda)$的全部初等因子。将$(\lambda-\lambda_1)$的指数

$$e_{11},\ e_{21},\cdots,e_{r_1 1},\ h_{11},\ h_{21},\ \cdots,\ h_{r_2 1}$$

按大小的顺序来排列，命为

$$c_1,\ c_2,\ \cdots,\ c_r$$

即

$$0\leqslant c_1\leqslant c_2\leqslant\cdots\leqslant c_r$$

因为$A(\lambda)$是由$B(\lambda)$与$C(\lambda)$所构成的准对角形矩阵，所以在矩阵$B(\lambda)$或$C(\lambda)$上施行的初等变换，实际上是在$A(\lambda)$上施行初等变换，于是

$$A(\lambda)\simeq\begin{pmatrix} d_1(\lambda) & & & & & & & & \\ & \ddots & & & & & & & \\ & & d_{r_1}(\lambda) & & & & & & \\ & & & \bar{d}_1(\lambda) & & & & & \\ & & & & \ddots & & & & \\ & & & & & \bar{d}_{r_2}(\lambda) & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{pmatrix}$$

$$
\simeq\begin{pmatrix}
(\lambda-\lambda_1)^{c_1}\varphi_1(\lambda) & & & & & & \\
 & (\lambda-\lambda_1)^{c_2}\varphi_2(\lambda) & & & & & \\
 & & \ddots & & & & \\
 & & & (\lambda-\lambda_1)^{c_r}\varphi_r(\lambda) & & & \\
 & & & & 0 & & \\
 & & & & & \ddots & \\
 & & & & & & 0
\end{pmatrix}
$$

式中 r 个多项式 $\varphi_1(\lambda)$，$\varphi_2(\lambda)$，…，$\varphi_r(\lambda)$ 都不含因式 $(\lambda-\lambda_1)$。设 $A(\lambda)$的行列式因子为

$$D_1^*(\lambda),\ D_2^*(\lambda),\ \cdots,\ D_r^*(\lambda)$$

所以在这些行列式因子中因式 $\lambda-\lambda_1$ 的最高幂指数分别等于

$$c_1,\ \sum_{i=1}^{2}c_i,\ \cdots,\ \sum_{i=1}^{r-1}c_i,\ \sum_{i=1}^{r}c_i$$

根据行列因子与不变因子的关系式（2-5），则知含在不变因子 $d_1^*(\lambda)$，$d_2^*(\lambda)$，…，$d_r^*(\lambda)$ 中因式 $\lambda-\lambda_1$ 的最高幂指数分别等于 c_1，c_2，…c_r，这就是说，$A(\lambda)$ 中与 $\lambda-\lambda_1$ 相应的初等因子是

$$(\lambda-\lambda_1)^{c_1},\ (\lambda-\lambda_1)^{c_2},\ \cdots(\lambda-\lambda_1)^{c_r}$$

中不为零指数的幂 $(\lambda-\lambda_1)^{c_i}$（即 $c_i\neq 0$），因而就是 $B(\lambda)$、$C(\lambda)$ 中与 $\lambda-\lambda_1$ 相应的全部初等因子。同理，对 $\lambda-\lambda_2$，$\lambda-\lambda_3$，…$\lambda-\lambda_t$ 也得相同结论。于是我们证明了 $B(\lambda)$、$C(\lambda)$的全部初等因子都是 $A(\lambda)$的初等因子。

剩下要证明，除此之外，$A(\lambda)$再没有别的初等因子。

设 $(\lambda-a)^k$ 是 $A(\lambda)$ 的一个初等因子，于是 $(\lambda-a)^k$ 一定是包含在某一个不变因子 $d_i^*(\lambda)$中 $\lambda-a$ 的最高次幂，因此，$(\lambda-a)^k|d_r^*(\lambda)$，故 $(\lambda-a)^k|D_r^*(\lambda)$，此即 $\lambda=a$ 是 $D_r^*(\lambda)$ 的一个根，即 $D_r^*(a)=0$。另一方面，由于

$$
A(\lambda)\simeq\begin{pmatrix}
d_1(\lambda) & & & & & & & & \\
 & \ddots & & & & & & & \\
 & & d_{r_1}(\lambda) & & & & & & \\
 & & & \tilde{d}_1(\lambda) & & & & & \\
 & & & & \ddots & & & & \\
 & & & & & \tilde{d}_{r_2}(\lambda) & & & \\
 & & & & & & 0 & & \\
 & & & & & & & \ddots & \\
 & & & & & & & & 0
\end{pmatrix}
$$

故

$$D_r^*(\lambda)\equiv d_1(\lambda)\cdots d_{r_1}(\lambda)\tilde{d}_1(\lambda)\cdots\tilde{d}_{r_2}(\lambda)$$

因为

$$d_i(\lambda)\,|\,d_{r_1}(\lambda),\qquad \tilde{d}_j(\lambda)\,|\,\tilde{d}_{r_2}(\lambda),$$

$$(\,i=1,\ 2,\ \cdots,\ r_1;\ j=1,\ 2,\ \cdots r_2\,)$$

所以

$$\tilde{d}_{r_1}(a)\tilde{d}_{r_2}(a)=0$$

这表明 a 必是 λ_1，λ_2，…，λ_t 中的某一个，所以$(\lambda-a)^k$是与某个 $\lambda-\lambda_i$ 相应的一个初等因子，因此由上面证明可知，$(\lambda-a)^k$ 一定是某个 $(\lambda-a)^{h_{ij}}$ 或 $(\lambda-a)^{e_{ji}}$$(i=1,$ $2,\cdots,r_1;\ j=1,\ 2,\ \cdots,\ r_2)$，此即证明了除 $B(\lambda)$ 与 $C(\lambda)$ 的全部初等因子外，$A(\lambda)$ 再没有别的初等因子。 】

应用归纳法，可以很容易地证明：对于一般准对角矩阵有类似于定理 2.2.7 之结论，即

定理 2.2.8 若 λ-矩阵

$$A(\lambda)=\begin{pmatrix} B_1(\lambda) & & & \\ & B_2(\lambda) & & \\ & & \ddots & \\ & & & B_t(\lambda) \end{pmatrix}$$

则 $B_1(\lambda)$，$B_2(\lambda)$，…，$B_t(\lambda)$ 各个初等因子的全体构成 $A(\lambda)$ 的全部初等因子。

证略。

根据定理 2.2.8 立即得到下述结论。

定理 2.2.9 设λ-矩阵

$$A(\lambda)=\begin{pmatrix} f_1(\lambda) & & & & & \\ & f_2(\lambda) & & & & \\ & & \ddots & & & \\ & & & f_r(\lambda) & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix}$$

则 $f_1(\lambda)$，$f_2(\lambda)$，…，$f_r(\lambda)$ 的所有一次因式的幂积构成 $A(\lambda)$ 的全部初等因子。

例 2.2.1 求 λ-矩阵

$$A(\lambda)=\begin{pmatrix} \lambda-a & b_1 & & & 0 \\ & \lambda-a & b_2 & & \\ & & \lambda-a & \ddots & \\ & & & \ddots & b_{n-1} \\ 0 & & & & \lambda-a \end{pmatrix}$$

的不变因子和初等因子，其中 b_1，b_2，…，b_{n-1} 都是常数，$b_1b_2\cdots b_{n-1}\neq0$。

〔解〕 $A(\lambda)$的行列式因子易得为

$$D_1(\lambda)=D_2(\lambda)=\cdots=D_{n-1}(\lambda)=1,\ D_n(\lambda)=(\lambda-a)^n$$

于是 $A(\lambda)$ 的不变因子为

$$d_1(\lambda)=d_2(\lambda)=\cdots=d_{n-1}(\lambda)=1,\ d_n(\lambda)=(\lambda-a)^n$$

因而初等因子只有一个

$$(\lambda-a)^n$$

例 2.2.2 求λ-矩阵

$$A(\lambda)=\begin{pmatrix} \lambda & & & a_n \\ -1 & \lambda & 0 & a_{n-1} \\ & -1 & \ddots & \vdots \\ & & \ddots \ \lambda & a_2 \\ 0 & & -1 & \lambda+a_1 \end{pmatrix}$$

的标准形。

〔解〕 将 $A(\lambda)$ 之第二行，第三行，…，第 n 行分别乘以 λ，λ^2，…，λ^{n-1} 都加到第一行上，得到

$$A(\lambda)=\begin{pmatrix} 0 & & & f(\lambda) \\ -1 & \lambda & 0 & a_{n-1} \\ & -1 & \ddots & \vdots \\ & & \ddots \ \lambda & a_2 \\ & 0 & -1 & \lambda+a_1 \end{pmatrix}$$

其中

$$f(\lambda)=\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-2}\lambda^2+a_{n-1}\lambda+a_n$$

易得 $\det A(\lambda)=f(\lambda)$

故 $$D_n(\lambda)=f(\lambda)$$

又 $$D_{n-1}(\lambda)=1$$

于是

$$D_1(\lambda)=D_2(\lambda)=\cdots=D_{n-2}(\lambda)=1$$

所以

$$d_1(\lambda)=d_2(\lambda)=\cdots=d_{n-1}(\lambda)=1,\ d_n(\lambda)=f(\lambda)$$

因此 $A(\lambda)$ 之标准形为

$$A(\lambda)\simeq\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & f(\lambda) \end{pmatrix}$$

§2.3 λ-矩阵的除法

为了讨论 λ-矩阵的除法，我们先来给出 λ-矩阵的多项式表示。

设 $m\times n$ λ-矩阵

$$A(\lambda)=\begin{pmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ a_{21}(\lambda) & a_{22}(\lambda) & \cdots & a_{2n}(\lambda) \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots & a_{mn}(\lambda) \end{pmatrix}$$

称 mn 个多项式 $a_{ij}(\lambda)$ 中次数最高的数 l 为 $A(\lambda)$的次，即

$$l=\max_{\substack{1\le i\le m\\ 1\le j\le n}}\deg a_{ij}(x)$$

其中 $\deg a_{ij}(\lambda)$ 表示多项式 $a_{ij}(\lambda)$ 的次数。元素 $a_{ij}(\lambda)$可以表示成

$$a_{ij}(\lambda)=\alpha_{ij}^{(0)}+\alpha_{ij}^{(1)}\lambda+\cdots+\alpha_{ij}^{(l)}\lambda^l$$

$(i=1,\ 2,\ \cdots,\ m;\ j=1,\ 2,\ \cdots,\ n)$，并且 mn 个 $\alpha_{ij}^{(l)}$ 中至少有一个不为零。

命

$$A_r=\begin{pmatrix} \alpha_{11}^{(r)} & \alpha_{12}^{(r)} & \cdots & \alpha_{1n}^{(r)} \\ \alpha_{21}^{(r)} & \alpha_{22}^{(r)} & \cdots & \alpha_{2n}^{(r)} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{m1}^{(r)} & \alpha_{m2}^{(r)} & \cdots & \alpha_{mn}^{(r)} \end{pmatrix}\qquad (r=0,\ 1,\ 2,\ \cdots,\ l)$$

显然 A_l 不是零矩阵，于是 $A(\lambda)$ 可以表示为

$$A(\lambda)=A_0+A_1\lambda+\cdots+A_l\lambda^l \tag{2-8}$$

这是一个以常数矩阵为系数的λ的多项式。这就是说，多项式矩阵可以表示成一个以常数矩阵为系数的多项式，A_l 称为 $A(\lambda)$ 的**首项系数**，如果 $A(\lambda)$ 是 n 阶矩阵，并且 $A_l=\pm I_n$，则称 $A(\lambda)$ 的首项系数为 1 。式（2-8）称为 $A(\lambda)$ 的**多项式表示**。

例 2.3.1 设

$$A(\lambda)=\begin{pmatrix}\lambda^2+\lambda+1 & \lambda^2-\lambda+2\\ 2\lambda & \lambda^2-3\lambda-1\end{pmatrix}$$

则 $A(\lambda)$ 的多项式表示为

$$A(\lambda)=\begin{pmatrix}1 & 2\\ 0 & -1\end{pmatrix}+\begin{pmatrix}1 & -1\\ 2 & -3\end{pmatrix}\lambda+\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\lambda^2$$

设 $A(\lambda)$、$B(\lambda)$ 分别为 l 次和 m 次 n 阶λ-矩阵，即

$$A(\lambda)=\sum_{i=0}^{l}A_i\lambda^i,\qquad B(\lambda)=\sum_{i=1}^{m}B_i\lambda^i$$

则

$$A(\lambda)+B(\lambda)\overset{\Delta}{=}\sum_{i=0}^{h}(A_i+B_i)\lambda^i \tag{2-9}$$

其中

$$h=\max(l,\ m)$$

$$A(\lambda)B(\lambda)\overset{\Delta}{=}A_0B_0+\lambda(A_1B_0+A_0B_1)+\cdots+\lambda^{l+m}A_lB_m \tag{2-10}$$

它的次数不大于 $l+m$ 。显然，若 $A(\lambda)$ 和 $B(\lambda)$ 的首项系数 A_l 和 B_m 都是可逆的，那么积 $A(\lambda)B(\lambda)$的次数恰好是 $l+m$ 。

定义 2.3.1 设 $A(\lambda)$、$B(\lambda)$ 分别为 l 次和 m 次 λ-矩阵，且 $B(\lambda)$的首项系数为可逆矩阵，若存在λ-矩阵$Q(\lambda)$、$R(\lambda)$，使得

$$A(\lambda)=Q(\lambda)B(\lambda)+R(\lambda) \tag{2-11}$$

其中 $R(\lambda)\equiv 0$ 或 $\deg R(\lambda)<m$ ，则称 $Q(\lambda)$ 是 $A(\lambda)$用 $B(\lambda)$除的**右商**，$R(\lambda)$是$A(\lambda)$ 用 $B(\lambda)$ 除的**右余**。类似地，可以定义**左商**和**左余**。

若 $A(\lambda)$ 用 $B(\lambda)$ 除的右余是零，那么 $Q(\lambda)$ 就叫做 $A(\lambda)$ 用 $B(\lambda)$ 除的**右因子**。类似地，可以定义**左因子**。

定理 2.3.1 设

$$A(\lambda)=\sum_{i=0}^{l}\lambda^iA_i \ \text{与}\ B(\lambda)=\sum_{i=0}^{m}\lambda^iB_i$$

分别是次数为 l 与 m 的 n 阶λ-矩阵，且 $\det B_m\neq 0$ ，那么存在$A(\lambda)$用$B(\lambda)$除的右商和右

余。同时还存在$A(\lambda)$用$B(\lambda)$除的左商和左余。

〔证明〕 若 $l<m$，则可取 $Q(\lambda)=0$，$R(\lambda)=A(\lambda)$ 便得所需之结果。

若 $l\geqslant m$，用 $B_m^{-1}\lambda^{-m}$ 左乘 $B(\lambda)$，得

$$B_m^{-1}\lambda^{-m}B(\lambda)=I+\lambda^{-1}B_m^{-1}B_{m-1}+\lambda^{-2}B_m^{-1}B_{m-2}$$
$$+\cdots+\lambda^{-m+1}B_m^{-1}B_1+\lambda^{-m}B_m^{-1}B_0$$

由上式可得

$$\lambda^l I=B_m^{-1}\lambda^{l-m}B(\lambda)-\lambda^{l-1}B_m^{-1}B_{m-1}-\lambda^{l-2}B_m^{-1}B_{m-2}$$
$$-\cdots-\lambda^{l-m+1}B_m^{-1}B_1-\lambda^{l-m}B_m^{-1}B_0$$

代入 $A(\lambda)$，得

$$\begin{aligned}A(\lambda)&=A_l\lambda^l+A_{l-1}\lambda^{l-1}+\cdots+A_1\lambda+A_0\\&=A_l(B_m^{-1}\lambda^{l-m}B(\lambda)-\lambda^{l-1}B_m^{-1}B_{m-1}-\cdots\\&\quad-\lambda^{l-m+1}B_m^{-1}B_1-\lambda^{l-m}B_m^{-1}B_0)\\&\quad+A_{l-1}\lambda^{l-1}+\cdots+A_1\lambda+A_0\\&=A_lB_m^{-1}\lambda^{l-m}B(\lambda)+A^{(1)}(\lambda)\end{aligned}\qquad(2\text{-}12)$$

其中$A^{(1)}(\lambda)$的次数$l_1\leqslant l-1$，且

$$A^{(1)}(\lambda)=A_{l_1}^{(1)}\lambda^{l_1}+\cdots+A_1^{(1)}\lambda+A_0^{(1)}$$

其中 $A_{l_1}^{(1)}\neq0$，$l_1<l$。若 $l_1\geqslant m$，类似于上述步骤，可得

$$A^{(2)}(\lambda)=A_{l_2}^{(2)}\lambda^{l_2}+\cdots+A_1^{(2)}\lambda+A_0^{(2)}$$

其中 $A_{l_2}^{(2)}\neq0$，$l_2<l_1$。

继续这样下去，我们可以得到一组 λ-矩阵

$$A(\lambda),\ A^{(1)}(\lambda),\ A^{(2)}(\lambda),\ \cdots,\ A^{(r)}(\lambda)$$

它们的次数越来越低，所以一定有 $A^{(r)}(\lambda)$，它的次数 $l_r<m$，而 $l_{r-1}\geqslant m$。若记$A(\lambda)=A^{(0)}(\lambda)$，则这组 λ-矩阵的每一个 $A^{(s-1)}(\lambda)$ 都有

$$A^{(s-1)}(\lambda)=A_{l_{s-1}}^{(s-1)}B_m^{-1}\lambda^{l_{s-1}-m}B(\lambda)+A^{(s)}(\lambda),$$
$$(s=1,\ 2,\ \cdots,\ r)\qquad(2\text{-}13)$$

最后得到

$$A(\lambda)=(A_lB_m^{-1}\lambda^{l-m}+A_{l_1}^{(1)}B_m^{-1}\lambda^{l_1-m}+\cdots$$
$$+A_{l_{r-1}}^{(r-1)}B_m^{-1}\lambda^{l_{r-1}-m})B(\lambda)+A^{(r)}(\lambda)\qquad(2\text{-}14)$$

这表明 $A(\lambda)$ 用$B(\lambda)$除的右商是上式右端括号中的部分，右余是$A^r(\lambda)$。

类似地，可证左商与左余的存在。 ■

定理 2.3.2 定理 2.3.1 中的右（左）商与右（左）余都是唯一的。

〔证明〕 设存在 λ-矩阵 $Q(\lambda)$，$R(\lambda)$ 和 $Q_1(\lambda)$、$R_1(\lambda)$ 使得

$$A(\lambda)=Q(\lambda)B(\lambda)+R(\lambda)$$

和

$$A(\lambda)=Q_1(\lambda)B(\lambda)+R_1(\lambda)$$

其中 $R(\lambda)$ 和 $R_1(\lambda)$ 的次数都小于 m，那么

$$(Q(\lambda)-Q_1(\lambda))B(\lambda)=R(\lambda)-R_1(\lambda)$$

若 $Q(\lambda)-Q_1(\lambda)\neq 0$，那么上式左边的次数最小为 m($B(\lambda)$的次数)，而右边的次数 $<m$，矛盾！因此 $Q_1(\lambda)=Q(\lambda)$，于是 $R_1(\lambda)=R(\lambda)$。

类似地，可以证明左商与左余的唯一性。 ▮

例 2.3.1 设

$$A(\lambda)=\begin{pmatrix}\lambda^4+\lambda^2+\lambda-1 & \lambda^3+\lambda^2+\lambda+2\\ 2\lambda^3-\lambda & 2\lambda^2+2\lambda\end{pmatrix}$$

和

$$B(\lambda)=\begin{pmatrix}\lambda^2+1 & 1\\ \lambda & \lambda^2+\lambda\end{pmatrix}$$

试求 $A(\lambda)$ 用 $B(\lambda)$ 除的右（左）商与右（左）余。

〔解〕 因为

$$B(\lambda)=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\lambda^2+\begin{pmatrix}0 & 0\\ 1 & 1\end{pmatrix}\lambda+\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}$$

所以 $B(\lambda)$ 的首项系数为 I，可逆，又由式（2-14）得

$$A(\lambda)=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\lambda^4+\begin{pmatrix}0 & 1\\ 2 & 0\end{pmatrix}\lambda^3+\begin{pmatrix}1 & 1\\ 0 & 2\end{pmatrix}\lambda^2$$

$$+\begin{pmatrix}1 & 1\\ -1 & 2\end{pmatrix}\lambda+\begin{pmatrix}-1 & 2\\ 0 & 0\end{pmatrix}$$

$$=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\lambda^2 B(\lambda)+A^{(1)}(\lambda)$$

不难知

$$A^{(1)}(\lambda)=\begin{pmatrix}0 & 1\\ 2 & 0\end{pmatrix}\lambda^3+\begin{pmatrix}0 & 0\\ 0 & 2\end{pmatrix}\lambda^2+\begin{pmatrix}1 & 1\\ -1 & 2\end{pmatrix}\lambda$$

$$+\begin{pmatrix}-1 & 2\\ 0 & 0\end{pmatrix}$$

又由式（2-13）可得

$$A^{(1)}(\lambda)=\begin{pmatrix}0 & 1\\ 2 & 0\end{pmatrix}\lambda B(\lambda)+A^{(2)}(\lambda)$$

其中

$$A^{(2)}(\lambda)=\begin{pmatrix}-1 & -1\\ 0 & 2\end{pmatrix}\lambda^2+\begin{pmatrix}1 & 1\\ -3 & 0\end{pmatrix}\lambda+\begin{pmatrix}-1 & 2\\ 0 & 0\end{pmatrix}$$

类似地

$$A^{(2)}(\lambda)=\begin{pmatrix}-1 & -1\\ 0 & 2\end{pmatrix}B(\lambda)+A^{(3)}(\lambda)$$

其中

$$A^{(3)}(\lambda)=\begin{pmatrix}2 & 2\\ -5 & -2\end{pmatrix}\lambda+\begin{pmatrix}0 & 3\\ 0 & 0\end{pmatrix}$$

因此

$$A(\lambda)=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\lambda^2 B(\lambda)+\begin{pmatrix}0 & 1\\ 2 & 0\end{pmatrix}\lambda B(\lambda)$$

$$+\begin{pmatrix}-1 & -1\\ 0 & 2\end{pmatrix}B(\lambda)+\begin{pmatrix}2 & 2\\ -5 & -2\end{pmatrix}\lambda+\begin{pmatrix}0 & 3\\ 0 & 0\end{pmatrix}$$

$$=\begin{pmatrix}\lambda^2-1 & \lambda-1\\ 2\lambda & 2\end{pmatrix}B(\lambda)+\begin{pmatrix}2\lambda & 2\lambda+3\\ -5\lambda & -2\lambda\end{pmatrix}$$

故 $A(\lambda)$ 用 $B(\lambda)$ 除的右商

$$Q(\lambda)=\begin{pmatrix}\lambda^2-1 & \lambda-1\\ 2\lambda & 2\end{pmatrix}$$

右余

$$R(\lambda)=\begin{pmatrix}2\lambda & 2\lambda+3\\ -5\lambda & -2\lambda\end{pmatrix}$$

类似地，可得

$$A(\lambda)=B(\lambda)\begin{pmatrix}\lambda^2 & \lambda+1\\ \lambda-1 & 1\end{pmatrix}$$

此即 $A(\lambda)$ 用 $B(\lambda)$ 除的左因子为

$$\begin{pmatrix} \lambda^2 & \lambda+1 \\ \lambda-1 & 1 \end{pmatrix}$$

现在我们来讨论一种重要的特殊情况。

设$A(\lambda)$为n阶l次λ-矩阵

$$A(\lambda)=A_l\lambda^l+A_{l-1}\lambda^{l-1}+\cdots+A_1\lambda+A_0$$
$$=\lambda^l A_l+\lambda^{l-1}A_{l-1}+\cdots+\lambda A_1+A_0$$

B是n阶数字矩阵，我们定义

$$A(B)\triangleq A_l B^l+A_{l-1}B^{l-1}+\cdots+A_1 B+A_0$$
$$\tilde{A}(B)\triangleq B^l A_l+B^{l-1}A_{l-1}+\cdots+BA_1+A_0$$

称$A(B)$是$A(\lambda)$在B处的右值，称$\tilde{A}(B)$是$A(\lambda)$在B处的左值。

由纯量多项式的余数定理知，纯量多项式$p(\lambda)$用$\lambda-b$除的余式为$p(b)$，对λ-矩阵也有类似的结论。

定理 2.3.3 λ-矩阵$A(\lambda)$用$\lambda I-B$除的右余和左余分别为$A(B)$和$\tilde{A}(B)$。

〔证明〕 不难验证

$$\lambda^j I-B^j=(\lambda^{j-1}I+\lambda^{j-2}B+\cdots+\lambda B^{j-2}$$
$$+B^{j-1})(\lambda I-B)$$

上式两边左乘A_j得l个方程

$$A_j(\lambda^j I-B^j)=A_j(\lambda^{j-1}I+\lambda^{j-2}B+\cdots$$
$$+\lambda B^{j-2}+B^{j-1})(\lambda I-B)$$
$$(j=1,\ 2,\ \cdots,\ l)$$

把这l个方程左、右两边分别相加，左式相加得

$$\sum_{j=1}^{l}A_j(\lambda^j I-B^j)=\sum_{j=1}^{l}A_j\lambda^j-\sum_{j=1}^{l}A_jB^j$$
$$=\sum_{j=0}^{l}A_j\lambda^j-\sum_{j=0}^{l}A_jB^j=A(\lambda)-A(B)$$

右式相加得

$$\sum_{j=1}^{l}A_j(\lambda^{j-1}I+\lambda^{j-2}B+\cdots+B^{j-1})(\lambda I-B)$$
$$=C(\lambda)(\lambda I-B)$$

其中$C(\lambda)$是次数$\leqslant l-1$的λ-矩阵，所以

$$A(\lambda)=C(\lambda)(\lambda I-B)+A(B)$$

这表明 $A(\lambda)$ 用 $\lambda I-B$ 除的右余为 $A(B)$。
类似地可得 $A(\lambda)$ 用 $\lambda I-B$ 除的左余为$\bar{A}(B)$。 ▮

定义 2.3.2 一个 n 阶常数矩阵 X 若满足 $A(X)=0$（或者 $\bar{A}(X)=0$）便称 X 是 λ-矩阵的**右解**（或者**左解**）。

根据定理 2.3.3 立即可得下述推论。

推论 2.3.1 λ-矩阵$A(\lambda)$用 $\lambda I-B$ 除是右（或左）除尽的充要条件为：B 是$A(\lambda)$的右解（或左解）。

作为推论 2.3.1 的一个例子，可以给出下述 Cayley-Hamilton 定理的一个证明。

定理 2.3.4 设 A 是数域 P 上一个 n 阶矩阵，$D(\lambda)=|\lambda I-A|$是A 的特征多项式，则 $D(A)=0$。

〔证明〕 设 $B(\lambda)=\mathrm{adj}(\lambda I-A)$

故

$$(\lambda I-A)B(\lambda)=B(\lambda)(\lambda I-A)=D(\lambda)I$$

因为 $D(\lambda)I$ 是 n 次 λ-矩阵，上式表明 $D(\lambda)I$ 用 $\lambda I-A$ 除的右余与左余都是零。由推论 2.3.1 便得

$$D(A)=\bar{D}(A)=0$$ ▮

推论 2.3.2 设 f 是数域 F 上的纯量多项式，A 是数域 F 上的一个 n 阶矩阵，那么存在一个依赖于 A 的次数小于 n 的纯量多项式 p，满足

$$f(A)=p(A)$$

〔证明〕 设

$$f(\lambda)=q(\lambda)D(\lambda)+r(\lambda)$$

其中 $D(\lambda)$ 是 A 的特征多项式，$r(\lambda)$ 或为零或它的次数不大于 $n-1$ 的多项式。因此

$$f(A)=q(A)D(A)+r(A)$$

由定理 2.3.4 知$D(A)=0$，所以

$$f(A)=r(A)$$ ▮

§2.4 矩阵相似的条件

我们知道，数字矩阵的特征矩阵是一种特殊的 λ-矩阵，它是研究数字矩阵的重要工具。在下面的讨论中将会看到，数字矩阵的相似可以归结为特征矩阵的等价。

定理 2.4.1 设 A、B 是两个 n 阶数字矩阵，则$A\sim B$ 的充要条件是$\lambda I-A\sim\lambda I-B$。

〔证明〕 *必要性* 设 $A\sim B$，则存在满秩矩阵 P，使得

$$P^{-1}AP=B$$

故

$$\lambda I-B=\lambda I-P^{-1}AP=P^{-1}(\lambda I-A)P$$

充分性　设

$$P^{-1}(\lambda I-A)P=\lambda I-B$$

故

$$\lambda I-P^{-1}AP=\lambda I-B$$

即

$$B=P^{-1}AP$$

∎

定理2.4.2　$A\sim B$的充要条件是 $\lambda I-A$ 与 $\lambda I-B$ 等价。

〔证明〕　由定理2.4.1知必要性显然。现证充分性。设 $\lambda I-A$ 与 $\lambda I-B$ 等价，则存在可逆 λ-矩阵$U(\lambda)$、$V(\lambda)$，使得

$$\lambda I-A=U(\lambda)(\lambda I-B)V(\lambda)$$

或

$$U^{-1}(\lambda)(\lambda I-A)=(\lambda I-B)V(\lambda) \tag{2-15}$$

由定理2.3.1知

$$U(\lambda)=(\lambda I-A)Q(\lambda)+U_0 \tag{2-16}$$

$$V(\lambda)=R(\lambda)(\lambda I-A)+V_0 \tag{2-17}$$

其中U_0、V_0为常数矩阵。把式(2-16)代入式(2-15)得

$$U^{-1}(\lambda)(\lambda I-A)=(\lambda I-B)R(\lambda)(\lambda I-A)+(\lambda I-B)V_0$$

或

$$[U^{-1}(\lambda)-(\lambda I-B)R(\lambda)](\lambda I-A)=(\lambda I-B)V_0 \tag{2-18}$$

上式右端是一个次数为1的 λ-矩阵，而左端 $\lambda I-A$ 也是一个次数为1的 λ-矩阵，所以

$$U^{-1}(\lambda)-(\lambda I-B)R(\lambda)=P \tag{2-19}$$

必是一个数字矩阵。因此式(2-18)可以写成

$$P(\lambda I-A)=(\lambda I-B)V_0 \tag{2-20}$$

现在证明 P 可逆，且$P=U_0^{-1}$。

由式(2-19)可得

$$U(\lambda)P=I-U(\lambda)(\lambda I-B)R(\lambda)$$

即

$$I=U(\lambda)P+U(\lambda)(\lambda I-B)R(\lambda)$$

由式（2-15）可把上式改为

$$I=U(\lambda)P+(\lambda I-A)V^{-1}(\lambda)R(\lambda)$$

把式（2-16）代入上式便得

$$\begin{aligned}I&=[(\lambda I-A)Q(\lambda)+U_0]P+(\lambda I-A)V^{-1}(\lambda)R(\lambda)\\&=U_0P+(\lambda I-A)[Q(\lambda)P+V^{-1}(\lambda)R(\lambda)]\end{aligned}$$

比较上式两边λ-矩阵的次数知，上式右边第二项必为零，故

$$I=U_0P$$

即P可逆，且

$$P=U_0^{-1} \tag{2-21}$$

把式（2-21）代入式（2-19）得

$$U_0^{-1}(\lambda I-A)=(\lambda I-B)V_0$$

或

$$\begin{aligned}\lambda I-A&=U_0(\lambda I-B)V_0\\&=\lambda U_0V_0-U_0BV_0\end{aligned}$$

比较上式两端得

$$U_0V_0=I,\quad A=U_0BV_0$$

故得

$$U_0=V_0^{-1},\ A=V_0^{-1}BV_0 \qquad \blacksquare$$

为了以后需要，在这节最后我们介绍相似矩阵的某些性质。

设$A\sim B$，则

（1） $\mathrm{rank}A=\mathrm{rank}B$

（2） $A^T\sim B^T$

（3） $A^k\sim B^k$

（4） $p(A)\sim p(B)$，其中$p(\lambda)$为某一个纯量多项式。

证明作为练习留给读者。

§2.5 矩阵的有理标准形

从第一章的讨论中我们看到，不是任何数字矩阵都能与对角矩阵相似。因此，自然会提出这样的问题：如果矩阵A不能与对角矩阵相似，A能与什么简单形式的矩阵相似？这就是下边要讨论的问题。

由定理2.4.2知，A与B相似的充要条件是$\lambda I-A$与$\lambda I-B$等价。因此，自然要考虑

$\lambda I-A$ 与 $\lambda I-B$ 的不变因子和初等因子。

定理 2.5.1 A 与 B 相似的充要条件是 A 与 B 有相同的不变因子。

〔证明〕 由定理 2.4.2 知，$A\sim B$ 当且仅当 $\lambda I-A\simeq\lambda I-B$。又由定理 2.2.5 知定理为真。]

定理 2.5.2 A 与 B 相似的充要条件是 A 与 B 有相同的初等因子。

〔证明〕 因为 $\lambda I-A$ 与 $\lambda I-B$ 的秩都是 n，故由定理 2.4.2 与定理 2.2.6 便得所需之结论。]

由例 2.2.2 知，n 阶数字矩阵

$$\begin{pmatrix} 0 & \cdots & 0 & -a_n \\ 1 & & & -a_{n-1} \\ & \ddots & & \vdots \\ & & 1 & -a_1 \end{pmatrix}$$

的不变因子为

$$\underbrace{1,\ 1,\ \cdots,\ 1,}_{(n-1)\text{个}}\qquad f(\lambda)=\lambda^n+a_1\lambda^{n-1}+\cdots+a_n$$

因此，给了一个 n 次多项式 $f(\lambda)$，则有一个 n 阶矩阵，它只有一个非常数 $f(\lambda)$ 作为它的不变因子。这样，就可以从不变因子着手去求与已知矩阵相似而形状又简单的矩阵。

设 A 为 n 阶常数矩阵，它的特性矩阵 $\lambda I-A$ 中非常数的不变因子有 k 个：$\varphi_1(\lambda)$，$\varphi_2(\lambda)$，…，$\varphi_k(\lambda)$。不妨设

$$\varphi_i(\lambda)=\lambda^{n_i}+a_{i1}\lambda^{n_i-1}+\cdots+a_{i,n_i-1}\lambda+a_{in_i}$$

$$(i=1,\ 2,\ \cdots,\ k)$$

因为 $\det(\lambda I-A)$ 是 λ 的 n 次多项式，所以 $n_1+n_2+\cdots+n_k=n$。并且 $\lambda I-A$ 的常数不变因子有 $(n-k)$ 个。命

$$C_i=\begin{pmatrix} 0 & \cdots & 0 & -a_{in_i} \\ 1 & & & -a_{i,n_i-1} \\ & \ddots & & \vdots \\ & & 1 & -a_{i1} \end{pmatrix} \tag{2-22}$$

显然，C_i 的不变因子为

$$\underbrace{1,\ \cdots,\ 1,}_{(n_i-1)\text{个}}\ \varphi_i(\lambda)$$

作 n 阶矩阵

$$F=\begin{pmatrix} C_1 & & & \\ & C_2 & & \\ & & \ddots & \\ & & & C_k \end{pmatrix} \tag{2-23}$$

则我们有

$$\lambda I-F=\begin{pmatrix} \lambda I_1-C_1 & & & \\ & \lambda I_2-C_2 & & \\ & & \ddots & \\ & & & \lambda I_k-C_k \end{pmatrix}$$

式中 I 是 n 阶单位矩阵，I_i 是 n_i 阶单位矩阵。因为 C_i 的不变因子是

$$\underbrace{1,\ 1,\ \cdots,\ 1}_{(n_i-1)\text{个}},\ \varphi_i(\lambda)$$

所以 λI_i-C_i 的标准形为

$$\lambda I_i-C_i\simeq\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \varphi_i(\lambda) \end{pmatrix}$$

因此

$$\lambda I-F\simeq\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \varphi_1(\lambda) & & & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & \varphi_k(\lambda) \end{pmatrix}$$

因为 $\varphi_i(\lambda)|\varphi_{i+1}(\lambda)$，所以 $\lambda I-F$ 的不变因子为

$$\underbrace{1,\ \cdots,\ 1}_{(n-k)\text{个}},\ \varphi_1(\lambda),\ \cdots,\ \varphi_k(\lambda)$$

这表明 $\lambda I-A$ 与 $\lambda I-F$ 有相同的标准形，于是

$$\lambda I-A\simeq\lambda I-F$$

因此，根据定理 2.4.2，可知

$$A\sim F$$

综上所述，我们有下面的定理。

定理 2.5.3 设 A 为 n 阶常数矩阵，它的非常数的不变因子为：$\varphi_1(\lambda)$，$\varphi_2(\lambda)$，…，$\varphi_k(\lambda)$，且设

$$\varphi_i(\lambda)=\lambda^{n_i}+a_{i1}\lambda^{n_i-1}+\cdots+a_{i\,n_i-1}\lambda+a_{i\,n_i}$$

（$i=1,\ 2,\ \cdots,\ k$）其中，$n_1+n_2+\cdots+n_k=n$。F 为准对角矩阵

$$F=\begin{pmatrix} C_1 & & & \\ & C_2 & & \\ & & \ddots & \\ & & & C_k \end{pmatrix}$$

其中 C_i 为形如式（2-22）的矩阵，则

$$A\sim F$$

定理 2.5.3 中的矩阵 F 称为 A 的**有理标准形**。

§2.6 矩阵的 Jordan 标准形

由例 2.2.1 知，n 阶数字矩阵

$$\begin{pmatrix} a & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & a \end{pmatrix}$$

的初等因子为 $(\lambda-a)^n$。因此给了一个一次因式的幂 $(\lambda-a)^n$，就有一个 n 阶矩阵，它的初等因子为 $(\lambda-a)^n$。这样，就可以从初等因子着手去求与已知矩阵相似而形状又简单的矩阵。

设 A 为 n 阶常数矩阵，它们特征矩阵 $\lambda I-A$ 的初等因子为

$$(\lambda-\lambda_1)^{n_1},\ (\lambda-\lambda_2)^{n_2},\ \cdots,\ (\lambda-\lambda_r)^{n_r}$$

其中 $n_1+n_2+\cdots+n_r=n$。作 n_i 阶数字矩阵

$$J_i=\begin{pmatrix} \lambda_i & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_i \end{pmatrix}_{n_i\times n_i} \tag{2-24}$$

命

$$J=\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{pmatrix} \tag{2-25}$$

则

$$A\sim J$$

事实上，因为

$$
\lambda I_i - J_i = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & (\lambda-\lambda_i)^{n_i} \end{pmatrix}
$$

所以

$$
\lambda I - J = \begin{pmatrix} \lambda I_1 - J_1 & & & \\ & \lambda I_2 - J_2 & & \\ & & \ddots & \\ & & & \lambda I_r - J_r \end{pmatrix}
$$

由定理 2.2.9 知，$\lambda I - J$ 的初等因子是

$$
(\lambda-\lambda_1)^{n_1},\ (\lambda-\lambda_2)^{n_2},\ \cdots,\ (\lambda-\lambda_r)^{n_r}
$$

由定理 2.5.2 知，$\lambda I - A \simeq \lambda I - J$。因此 $A \sim J$。

通常称 J 是 A 的 Jordan 标准形，每个 J_i 称为 Jordan 块。

下面我们讨论矩阵 A 的 Jordan 块与特征向量之间的关系。

由定理 2.4.1 知

$$
\lambda I - A \sim \lambda I - J
$$

且

$$
\operatorname{rank}(\lambda I - A) = \operatorname{rank}(\lambda I - J) = \sum_{i=1}^{r} (\lambda I_i - J_i)
$$

设 $\lambda_l (l=1, 2, \cdots, r)$ 是 A 的某一个特征值，则

$$
\lambda_l I_i - J_i = \begin{pmatrix} \lambda_l-\lambda_i & & & \\ 1 & \lambda_l-\lambda_i & & \\ & 1 & \ddots & \\ & & \ddots\ 1 & \lambda_l-\lambda_i \end{pmatrix}_{n_i \times n_i}
$$

显然，当 $\lambda_l \neq \lambda_i$ 时，

$$
\operatorname{rank}(\lambda_l I_i - J_i) = n_i
$$

当 $\lambda_l = \lambda_i$ 时，

$$
\operatorname{rank}(\lambda_l I_i - J_i) = n_i - 1
$$

因此

（1） 若 $\lambda_1, \lambda_2, \cdots, \lambda_r$ 全不相同，则

$$
\operatorname{rank}(\lambda_l I - A) = \sum_{i=1}^{r} \operatorname{rank}(\lambda_l I - J_i) = n - 1
$$

因此，线性齐次方程组

$$(\lambda_i I-A)x=0$$

的基础解系只含有一个解，于是 A 对应于 λ_i 的线性无关的特征向量只有一个，此即在 A 的 Jordan 标准形中，每一个 Jordan 块对应A有一个特征向量。这时，A 共有 r 个线性无关的特性向量。

（2） 若 λ_1，λ_2，…，λ_r 中有相同者，例如 $\lambda_1=\lambda_2$，即 J_1，J_2 的特征值相同，不难知

$$\text{rank}(\lambda_1 I-A)=n-2$$

所以A的特征值λ_1对应有两个线性无关的特征向量，因此，仍得到 A 的 Jordan 标准形中每一个Jordan 块对应 A 有一个特征向量的结论。

于是我们有下述定理。

定理 2.6.2 设A的 Jordan 标准形 J 由 r 个 Jordan 块 J_1，J_2，…，J_r 构成，则

（1） A 有 r 个线性无关的特征向量；

（2） 特征值 λ_i 的几何重复度 α_i 可由下式给出

$$\alpha_i=n-\text{rank}(\lambda_i I-A)$$

例 2.6.1 求

$$A=\begin{pmatrix}-1 & -2 & 6\\ -1 & 0 & 3\\ -1 & -1 & 4\end{pmatrix}$$

的 Jordan 标准形。

〔解〕 先求 A 的初等因子

$$\lambda I-A=\begin{pmatrix}\lambda+1 & 2 & -6\\ 1 & \lambda & -3\\ 1 & 1 & \lambda-4\end{pmatrix}\simeq\cdots\simeq\begin{pmatrix}1 & & \\ & \lambda-1 & \\ & & (\lambda-1)^2\end{pmatrix}$$

因此 A 的初等因子是

$$(\lambda-1),\ (\lambda-1)^2$$

故A的 Jordan 标准形是

$$J=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}$$

例 2.6.2 求解微分方程

$$\begin{cases} \dfrac{dx_1}{dt}=-x_1-2x_2+6x_3 \\ \dfrac{dx_2}{dt}=-x_1 \qquad +3x_3 \\ \dfrac{dx_3}{dt}=-x_1-x_2+4x_3 \end{cases} \tag{2-26}$$

〔解〕 设

$$\boldsymbol{x}=(x_1,\ x_2,\ x_3)^T,\quad \frac{d\boldsymbol{x}}{dt}=\left[\frac{dx_1}{dt},\ \frac{dx_2}{dt},\ \frac{dx_3}{dt}\right]^T$$

则方程组（2-26）可以写成

$$\frac{d\boldsymbol{x}}{dt}=A\boldsymbol{x}$$

其中

$$A=\begin{pmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{pmatrix}$$

由例 2.6.1 知

$$A\sim J=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

从而存在满秩矩阵 P，使

$$A=PJP^{-1}$$

于是

$$\frac{d\boldsymbol{x}}{dt}=PJP^{-1}\boldsymbol{x}$$

或
$$P^{-1}\frac{d\boldsymbol{x}}{dt}=JP^{-1}\boldsymbol{x}$$

令
$$\boldsymbol{y}=P^{-1}\boldsymbol{x}$$

于是得到方程

$$\frac{d\boldsymbol{y}}{dt}=J\boldsymbol{y}$$

把J代入上式，有

$$\frac{d\boldsymbol{y}}{dt}=\begin{pmatrix}1&0&0\\0&1&0\\0&1&1\end{pmatrix}\boldsymbol{y}$$

即

$$\frac{dy_1}{dt}=y_1$$

$$\frac{dy_2}{dt}=y_2$$

$$\frac{dy_3}{dt}=y_2+y_3$$

由此可以求出y_1、y_2和y_3。代入$\boldsymbol{x}=P\boldsymbol{y}$即可求出$x_1$、$x_2$、$x_3$。

由例2.6.2可以看到，在解决具体问题中不仅要求出Jordan标准形，而且需要求出变换矩阵P，下面我们来讨论矩阵P的求法。

设A的Jordan标准形为J，则

$$A\sim J=\begin{pmatrix}J_1&&&\\&J_2&&\\&&\ddots&\\&&&J_r\end{pmatrix}$$

其中

$$J_i=\begin{pmatrix}\lambda_i&&&\\1&\ddots&&\\&\ddots&\ddots&\\&&1&\lambda_i\end{pmatrix}_{n_i\times n_i}$$

把变换矩阵P按Jordan块J_i的阶数n_i进行相应的分块，即设

$$P=(P_1,\ P_2,\ \cdots,\ P_r)$$

其中P_i是$n\times n_i$阶矩阵。因此

$$\begin{aligned}&A(P_1,\ P_2,\ \cdots,\ P_r)\\&=(AP_1,\ AP_2,\ \cdots,\ AP_r)\end{aligned}$$

$$
=(P_1,\ P_2,\ \cdots,\ P_r)\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{pmatrix}
$$

比较上式两端得

$$
AP_i=P_iJ_i \qquad (i=1,\ 2,\ \cdots,\ r) \tag{2-27}
$$

命

$$
P_i=(x_{i1},\ x_{i2},\ \cdots,\ x_{in_i}) \tag{2-28}
$$

其中 x_{i1}，x_{i2}，…，x_{in_i} 均为 n 维列向量。显然，x_{i1}，x_{i2}，…，x_{in_i} 是线性无关的。把它代入式（2-27）得

$$
(Ax_{i1},\ Ax_{i2},\ \cdots,\ Ax_{in_i})
$$

$$
=(x_{i1},\ \cdots,\ x_{in_i})\begin{pmatrix} \lambda_i & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_i \end{pmatrix}
$$

比较上式两端便得下列方程组

$$
\begin{cases} Ax_{i1}=\lambda_i x_{i1}+x_{i2} \\ Ax_{i2}=\lambda_i x_{i2}+x_{i3} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ Ax_{in_i-1}=\lambda_i x_{in_i-1}+x_{in_i} \\ Ax_{in_i}=\lambda_i x_{in_i} \end{cases} \qquad (i=1,\ 2,\ \cdots,\ r) \tag{2-29}
$$

由式（2-29）的最后一个方程看到，列向量 x_{in_i} 是矩阵 A 的特征值为 λ_i 所对应的特征向量。且不难看到，由 x_{in_i} 可以求得 x_{in_i-1}，从而继续可以求得 x_{in_i-2}，…直至 x_{i1}。因此，长方形矩阵 P_i 可以求得。

由定理 2.6.2 知，满足式（2-29）的 x_{1n_1}，x_{2n_2}，… x_{rn_r} 是存在的，且是线性无关的。因此，对于每一个 P_i（$i=1,\ 2,\ \cdots,\ r$）都可以按上述步骤求得。从而可以求得 P。

由式（2-29）我们还可得到下面的定理。

定理 2.6.3 设 A 的 Jordan 标准形为

$$
J=\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{pmatrix}
$$

如果 J_1，J_2，…，J_r 中至少有一个是真 Jordan 块，则至少存在一个列向量 x，使得

$$
(\lambda I-A)x\neq 0
$$

但

$$(\lambda I-A)^p x=0$$

其中 $p\geqslant 2$ 为某一个整数。

〔证明〕 由式（2-29）知

$$\begin{cases}(\lambda_i I-A)x_{i1}=-x_{i2}\\(\lambda_i I-A)x_{i2}=-x_{i3}\\\cdots\cdots\cdots\cdots\cdots\cdots\\(\lambda_i I-A)x_{i n_i-1}=-x_{i n_i}\\(\lambda_i I-A)x_{i n_i}=0\end{cases}\tag{2-30}$$

依次以$(\lambda_i I-A)^{n_i-1}$，$(\lambda_i I-A)^{n_i-2}$，…，$(\lambda_i I-A)$，E 左乘方程组（2-30）的第一个方程，第二个方程，…，第 n_i-1 个方程，第 n_i 个方程得到

$$\begin{cases}(\lambda_i I-A)^{n_i}x_{i1}=0\\(\lambda_i I-A)^{n_i-1}x_{i2}=0\\\cdots\cdots\cdots\cdots\cdots\cdots\\(\lambda_i I-A)^2 x_{i n_i-1}=0\\(\lambda_i I-A)x_{i n_i}=0\end{cases}$$

因此只要 n_1，n_2，…，n_r 中至少有一个$\geqslant 2$（即 J_1，J_2，…，J_r 中至少有一个是 Jordan 块），那么满足定理中要求的向量x是存在的。 】

有时称满足定理中要求的向量x为A的广义特征向量。

习　题

2-1　用初等变换把下列 λ-矩阵化为标准形：

（1）$\begin{pmatrix}\lambda^3-\lambda & 2\lambda^2\\\lambda^2+5\lambda & 3\lambda\end{pmatrix}$　　（2）$\begin{pmatrix}\lambda^2-1 & 0\\0 & (\lambda-1)^3\end{pmatrix}$

（3）$\begin{pmatrix}1-\lambda & \lambda^2 & \lambda\\\lambda & \lambda & -\lambda\\1+\lambda^2 & \lambda^2 & -\lambda^2\end{pmatrix}$　　（4）$\begin{pmatrix}\lambda(\lambda+1) & 0 & 0\\0 & \lambda & 0\\0 & 0 & (\lambda+1)^2\end{pmatrix}$

（5）$\begin{pmatrix}0 & 0 & 0 & \lambda^2\\0 & 0 & \lambda^2-\lambda & 0\\0 & (\lambda-1)^2 & 0 & 0\\\lambda^2-\lambda & 0 & 0 & 0\end{pmatrix}$

2-2　不用初等变换把下列 λ-矩阵化为标准形：

（1）
$$\begin{pmatrix} \lambda(\lambda-1) & 0 & 0 \\ 0 & \lambda(\lambda-2) & 0 \\ 0 & 0 & (\lambda-1)(\lambda-2) \end{pmatrix}$$

（2）
$$\begin{pmatrix} \lambda-3 & -1 & 0 & 0 \\ 4 & \lambda+1 & 0 & 0 \\ -6 & -1 & \lambda-2 & -1 \\ 14 & 5 & 1 & \lambda \end{pmatrix}$$

（3）
$$\begin{pmatrix} \lambda+\alpha & \beta & 1 & 0 \\ -\beta & \lambda+\alpha & 0 & 1 \\ 0 & 0 & \lambda+\alpha & \beta \\ 0 & 0 & -\beta & \lambda+\alpha \end{pmatrix}$$

（4）
$$\begin{pmatrix} \lambda & 1 & 1 & \cdots\cdots & 1 \\ & \lambda & 1 & \cdots\cdots & 1 \\ & & \lambda & \cdots\cdots & 1 \\ & & & \ddots & \cdots\cdots \\ & & & & \lambda \end{pmatrix} \quad (n\text{阶})$$

2-3　求 λ-矩阵的标准形，如果已知它的初等因子，秩 r 以及阶 n 如下：

（1）　$\lambda+1$，$\lambda+1$，$(\lambda+1)^2$，$\lambda-1$，$(\lambda-1)^2$；$r=4$，$n=5$。

（2）　$\lambda+2$，$(\lambda+2)^2$，$(\lambda+2)^3$，$\lambda-2$，$(\lambda-2)^3$；$r=n=4$。

（3）　$\lambda-1$，$\lambda-1$，$(\lambda-1)^3$，$\lambda+2$，$(\lambda+2)^2$；$r=4$，$n=5$。

2-4　试证 Jordan 块

$$J(\lambda)=\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

相似于

$$\begin{pmatrix} \lambda & \varepsilon & 0 \\ 0 & \lambda & \varepsilon \\ 0 & 0 & \lambda \end{pmatrix}$$

这里 $\varepsilon\neq 0$ 是任意实数。

2-5 设 10 阶矩阵

$$A=\begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix},\quad B=\begin{pmatrix} \lambda & 1 & & & \\ 0 & \ddots & \ddots & & \\ \vdots & & \ddots & \ddots & \\ 0 & & & \ddots & 1 \\ \varepsilon & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

其中 $\varepsilon=10^{-10}$，证明B不相似于 A。

2-6 试证：$A\sim A'$。

2-7 设 $A\neq 0$，$A^k=0$， $(k\geqslant 2)$，则 A 不能与对角矩阵相似（A 称为幂零矩阵）。

2-8 设 $A^p=I$，则 A 与对角矩阵相似。

2-9 设 $A^2=A$，则

$$A\sim\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}$$

2-10 求下列矩阵的 Jordan 标准形和变换矩阵 P。

（1）$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ -2 & -2 & -1 \end{pmatrix}$ （2）$\begin{pmatrix} 13 & 16 & 16 \\ -5 & -7 & -6 \\ -6 & -8 & -7 \end{pmatrix}$

（3）$\begin{pmatrix} 4 & 5 & -2 \\ -2 & -2 & 1 \\ -1 & -1 & 1 \end{pmatrix}$ （4）$\begin{pmatrix} 1 & 1 & -1 \\ -3 & -3 & 3 \\ -2 & -2 & 2 \end{pmatrix}$

2-11 求下列矩阵的有理标准形

（1）$\begin{pmatrix} 1 & 1 & 1 \\ -3 & -3 & 3 \\ -2 & -2 & 2 \end{pmatrix}$ （2）$\begin{pmatrix} 4 & 5 & -2 \\ -2 & -2 & 1 \\ -1 & -1 & 1 \end{pmatrix}$

（4）$\begin{pmatrix} 3 & 1 & 0 & 0 \\ -4 & -1 & 0 & 0 \\ 6 & 1 & 2 & 1 \\ -14 & -5 & -1 & 0 \end{pmatrix}$ （4）$\begin{pmatrix} -2 & 2 & -1 & -1 \\ 2 & -2 & -1 & -1 \\ -1 & -1 & -2 & 2 \\ -1 & -1 & 2 & -2 \end{pmatrix}$

第三章　函数矩阵

前一章我们讨论了元素为 λ 的多项式的矩阵，这一章讨论更为一般的元素为 x 的实函数的矩阵。这种矩阵称之为函数矩阵，它是矩阵分析中所要研究的一个重要对象。

本章的内容有函数矩阵的概念及其基本性质，函数矩阵对纯量的导数和积分，函数向量的线性相关性。

§3.1　函数矩阵

以实变量 x 的函数为元素的矩阵

$$A(x)=\begin{pmatrix} a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x) \\ a_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x) \\ \cdots\cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ a_{m1}(x) & a_{m2}(x) & \cdots & a_{mn}(x) \end{pmatrix}_{m\times n}$$

称为函数矩阵，其中所有的元素 $a_{ij}(x)$（$i=1, 2, \cdots, m$；$j=1, 2, \cdots, n$）是定义在区间 $[a, b]$ 上的实函数。

当 $m=1$ 时，$A(x)$ 是一个函数行向量，当 $n=1$ 时，$A(x)$ 是一个函数列向量。

函数矩阵的加法、数乘、乘法和转置与常数矩阵的相应运算相同。即

加法　设 $A(x)=(a_{ij}(x))_{m\times n}$，$B(x)=(b_{ij}(x))_{m\times n}$，它们的和 $A(x)+B(x)$ 定义为

$$A(x)+B(x)\overset{\Delta}{=}(a_{ij}(x)+b_{ij}(x))_{m\times n}$$

数（量）乘　设 $k(x)$ 为 x 的函数，$A(x)=(a_{ij}(x))_{m\times n}$，$k(x)$ 与 $A(x)$ 的数乘积定义为

$$k(x)A(x)\overset{\Delta}{=}(k(x)a_{ij}(x))_{m\times n}$$

乘法　设 $A(x)=(a_{ij}(x))_{m\times r}$，$B(x)=(b_{ij}(x))_{r\times n}$，它们的积 $A(x)B(x)$ 定义为

$$A(x)B(x)\overset{\Delta}{=}(c_{ij}(x))_{m\times n}$$

其中 $c_{ij}(x)=\sum\limits_{t=1}^{r}a_{it}(x)b_{tj}(x)$　（$i=1, 2, \cdots, m$；$j=1, 2, \cdots, n$）。并说 $A(x)$ 和 $B(x)$ 是可乘的。

转置　设 $m\times n$ 阶矩阵

$$A(x)=\begin{pmatrix} a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x) \\ a_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x) \\ \cdots\cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ a_{m1}(x) & a_{m2}(x) & \cdots & a_{mn}(x) \end{pmatrix}_{m\times n}$$

的转置矩阵 $A^T(x)$ 为

$$A^T(x)=\begin{pmatrix} a_{11}(x) & a_{21}(x) & \cdots & a_{m1}(x) \\ a_{12}(x) & a_{22}(x) & \cdots & a_{m2}(x) \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ a_{1n}(x) & a_{2n}(x) & \cdots & a_{mn}(x) \end{pmatrix}_{n\times m}$$

由这些定义可知，它们的运算性质与常数矩阵的运算性质相同，这里不再列举。

定义 3.1.1 设 $A(x)=(a_{ij}(x))$ 为 n 阶函数矩阵，若存在 n 阶函数矩阵 $B(x)=(b_{ij}(x))$，使得对于任何 $x\in[a,\ b]$ 都有

$$A(x)B(x)\equiv B(x)A(x)\equiv I_n \tag{3-1}$$

则称 $A(x)$ 在 $[a,\ b]$ 上可逆，$B(x)$ 是 $A(x)$ 的逆矩阵，记为 $A^{-1}(x)$。

定理 3.1.1 n 阶矩阵 $A(x)$ 在 $[a,b]$ 上可逆的充要条件是 $|A(x)|$ 在 $[a,b]$ 上处处不为零，并且

$$A^{-1}(x)=\frac{1}{|A(x)|}\mathrm{adj}A(x)$$

其中

$$\mathrm{adj}A(x)=\begin{pmatrix} A_{11}(x) & A_{21}(x) & \cdots & A_{n1}(x) \\ A_{12}(x) & A_{22}(x) & \cdots & A_{n2}(x) \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots \\ A_{1n}(x) & A_{2n}(x) & \cdots & A_{nn}(x) \end{pmatrix}$$

$A_{ij}(x)$ 是 $A(x)$ 中元素 $a_{ij}(x)$ 的代数余子式。

证明作为练习留给读者。

定义 3.1.2 在 $[a,\ b]$ 上函数矩阵 $A(x)$ 不恒等于零的子式的最高阶数称为矩阵 $A(x)$ 的秩，记作 $\mathrm{rank}A(x)$。

定义 3.1.3 设 $A(x)$ 为 $[a,b]$ 上的 n 阶函数矩阵，如果 $\mathrm{rank}\ A(x)=n$，则称 $A(x)$ 为满秩的。

我们知道，对常数矩阵 A，如果 A 是可逆的，则 A 为满秩的；反之，如果 A 是满秩的，则 A 为可逆的。但是函数矩阵则不同，一个 n 阶函数矩阵 $A(x)$ 如果在 $[a,b]$ 上是可逆的，那么，行列式 $|A(x)|$ 在 $[a,b]$ 上处处不为零；因而 $A(x)$ 是满秩的。反之，一个 n 阶满秩函数矩阵却不一定是可逆的，这是因为当 $A(x)$ 是满秩时，只保证 $A(x)$ 的行列式不恒等于零，不排除 $|A(x)|$ 在 $[a,b]$ 上有零点，因此，$A(x)$ 在 $[a,b]$ 上不一定可逆。例如，

$$A(x)=\begin{pmatrix} x & 1 \\ 1 & x \end{pmatrix}$$

$|A(x)|=x^2-1$，所以在任何区间 $[a,b]$ 上，$A(x)$ 的秩为 2，它是满秩的。但是 $A(x)$ 在 $[a,b]$

上是否可逆还得由a,b的取值而定。当区间(a,b)包含±1时，$|A(x)|$在(a,b)上有零点，$A(x)$不可逆。只有在(a,b)不包含±1时，$|A(x)|$在(a,b)上无零点，$A(x)$才是可逆的。

常数矩阵A的特征矩阵$(\lambda I-A)$是含变量λ的矩阵，它是满秩的，在不包含A的特征值的区间上是可逆的，否则是不可逆的。

定义 3.1.4 若$A(x)=(a_{ij}(x))_{m\times n}$的所有各元素$a_{ij}(x)$在$x=x_0$处有极限，即

$$\lim_{x\to x_0} a_{ij}(x)=a_{ij} \qquad (i=1,\ 2,\ \cdots,\ m;\ j=1,\ 2,\ \cdots,\ n)$$

其中a_{ij}为固定常数。则称$A(x)$在$x=x_0$处有极限，且记为

$$\lim_{x\to x_0} A(x)=A$$

其中
$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

若$A(x)$的所有各元$a_{ij}(x)$在$x=x_0$处连续，即

$$\lim_{x\to x_0} a_{ij}(x)=a_{ij}(x_0) \qquad (i=1,\ 2,\ \cdots,m;\ j=1,\ 2,\ \cdots,\ n)$$

便称$A(x)$在$x=x_0$处连续，且记为

$$\lim_{x\to x_0} A(x)=A(x_0)$$

其中
$$A(x_0)=\begin{pmatrix} a_{11}(x_0) & a_{12}(x_0) & \cdots & a_{1n}(x_0) \\ a_{21}(x_0) & a_{22}(x_0) & \cdots & a_{2n}(x_0) \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}(x_0) & a_{m2}(x_0) & \cdots & a_{mn}(x_0) \end{pmatrix}$$

容易验证下列等式是成立的。

设 $\lim\limits_{x\to x_0} A(x)=A$, $\lim\limits_{x\to x} B(x)=B$

则

(1) $\lim\limits_{x\to x_0} [A(x)\pm B(x)]=A\pm B$

(2) $\lim\limits_{x\to x_0} [kA(x)]=kA$

(3) $\lim\limits_{x\to x} [A(x)B(x)]=AB$

§3.2 函数矩阵对纯量的导数与积分

定义 3.2.1 若 $A(x)=(a_{ij}(x))_{m\times n}$ 的所有各元 $a_{ij}(x)(i=1, 2, \cdots, m; \ j=1, 2, \cdots, n)$ 在点 $x=x_0$ 处（或在区间$[a, b]$上）可导，便称此函数矩阵 $A(x)$ 在点 $x=x_0$ 处（或在区间$[a, b]$上）可导，并且记为

$$
A'(x_0)=\left.\frac{\mathrm{d}A(x)}{\mathrm{d}x}\right|_{x=x_0}\triangleq\lim_{\Delta x\to 0}\frac{A(x_0+\Delta x)-A(x_0)}{\Delta x}
$$

$$
=\begin{pmatrix} a'_{11}(x_0) & a'_{12}(x_0) & \cdots & a'_{1n}(x_0) \\ a'_{21}(x_0) & a'_{22}(x_0) & \cdots & a'_{2n}(x_0) \\ \cdots & \cdots & \cdots & \cdots \\ a'_{m1}(x_0) & a'_{m2}(x_0) & \cdots & a'_{mn}(x_0) \end{pmatrix} \qquad (3\text{-}2)
$$

函数矩阵的导数运算有下列性质：

(1) $A(x)$ 是常数矩阵的充要条件是

$$\frac{\mathrm{d}A(x)}{\mathrm{d}x}=0$$

(2) 设 $A(x)=(a_{ij}(x))_{m\times n}$，$B(x)=(b_{ij}(x))_{m\times n}$均可导，则

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[A(x)+B(x)\right]=\frac{\mathrm{d}A(x)}{\mathrm{d}x}+\frac{\mathrm{d}B(x)}{\mathrm{d}x}$$

(3) 设 $k(x)$是 x 的纯量函数，$A(x)$是函数矩阵，$k(x)$与$A(x)$均可导，则

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[k(x)A(x)\right]=\frac{\mathrm{d}k(x)}{\mathrm{d}x}A(x)+k(x)\frac{\mathrm{d}A(x)}{\mathrm{d}x}$$

特别地，当 $k(x)$是常数 k 时有

$$\frac{\mathrm{d}}{\mathrm{d}x}[kA(x)]=k\frac{\mathrm{d}A(x)}{\mathrm{d}x}$$

(4) 设 $A(x)$、$B(x)$均可导，且 $A(x)$与 $B(x)$是可乘的，则

$$\frac{\mathrm{d}}{\mathrm{d}x}[A(x)B(x)]=\frac{\mathrm{d}A(x)}{\mathrm{d}x}B(x)+A(x)\frac{\mathrm{d}B(x)}{\mathrm{d}x}$$

因为矩阵乘法没有交换律，所以

$$\frac{\mathrm{d}}{\mathrm{d}x}A^2(x)\neq 2A(x)\frac{\mathrm{d}A(x)}{\mathrm{d}x}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}A^3(x)\neq 3A^2(x)\frac{\mathrm{d}A(x)}{\mathrm{d}x}$$

(5) 若 $A(x)$与 $A^{-1}(x)$都可导，则

$$\frac{dA^{-1}(x)}{dx}=-A^{-1}(x)\frac{dA(x)}{dx}A^{-1}(x) \tag{3-3}$$

〔证明〕 因为

$$A^{-1}(x)A(x)=I$$

所以

$$\frac{d}{dx}[A^{-1}(x)A(x)]=\frac{dA^{-1}(x)}{dx}A(x)+A^{-1}(x)\frac{dA(x)}{dx}=0$$

于是

$$\frac{dA^{-1}(x)}{dx}=-A^{-1}(x)\frac{dA(x)}{dx}A^{-1}(x)$$

】

(6) 设 $A(x)$ 为函数矩阵，$x=f(t)$是 t 的纯量函数，$A(x)$与 $f(t)$均可导，则

$$\frac{d}{dt}A(x)=\frac{dA(x)}{dx}f'(t)=f'(t)\frac{dA(x)}{dx}$$

函数矩阵的导数本身也是一个函数矩阵，它可以再进行求导运算。下面我们给出函数矩阵对纯量的高阶导数：

$$\frac{d^2A(x)}{dx}=\frac{d}{dx}\left(\frac{dA(x)}{dx}\right),\quad \frac{d^3A(x)}{dx}=\frac{d}{dx}\left(\frac{d^2A(x)}{dx}\right),$$

$$\cdots\cdots,\quad \frac{d^kA(x)}{dx}=\frac{d}{dx}\left(\frac{d^{k-1}A(x)}{dx}\right)$$

定义 3.2.2 若函数矩阵 $A(x)=(a_{ij}(x))_{m\times n}$ 的所有各元 $a_{ij}(x)$ $(i=1,2,\cdots,m;\ j=1,2,\cdots,n)$都在$[a,b]$上可积，则称 $A(x)$在$[a,b]$上可积，且

$$\int_a^b A(x)dx\triangleq\begin{pmatrix}\int_a^b a_{11}(x)dx & \int_a^b a_{12}(x)dx & \cdots & \int_a^b a_{1n}(x)dx\\ \cdots & \cdots & \cdots & \cdots\\ \int_a^b a_{m1}(x)dx & \int_a^b a_{m2}(x)dx & \cdots & \int_a^b a_{mn}(x)dx\end{pmatrix} \tag{3-4}$$

函数矩阵的定积分有如下简单性质：

(1) $$\int_a^b kA(x)dx=k\int_a^b A(x)dx \qquad k\in R$$

(2) $$\int_a^b [A(x)+B(x)]dx=\int_a^b A(x)dx+\int_a^b B(x)dx$$

函数矩阵的不定积分也可类似定义。

§3.3 函数向量的线性相关性

判断一个常数向量组是否线性相关的方法不适用函数向量。这节主要介绍两个判断函数向量相关性的方法。

定义 3.3.1 设有定义在$[a, b]$上的 m 个连续的函数向量

$$\boldsymbol{a}_i(x)=(a_{i1}(x), a_{i2}(x), \cdots, a_{in}(x))$$

$(i=1, 2, \cdots, m)$，若存在一组不全为零的实常数 $k_1, k_2, \cdots, k_m$，使得对于所有的 $x\in[a, b]$，等式

$$k_1\boldsymbol{a}_1(x)+k_2\boldsymbol{a}_2(x)+\cdots+k_m\boldsymbol{a}_m(x)=\boldsymbol{0} \tag{3-5}$$

成立，我们便说，在$[a, b]$上$\boldsymbol{a}_1(x), \boldsymbol{a}_2(x), \cdots, \boldsymbol{a}_m(x)$线性相关。否则就说 $\boldsymbol{a}_1(x), \boldsymbol{a}_2(x), \cdots, \boldsymbol{a}_m(x)$线性无关。即如果只有在 $k_1=k_2=\cdots=k_m=0$ 时(3-5)式才成立，那么就说 $\boldsymbol{a}_1(x), \boldsymbol{a}_2(x), \cdots, \boldsymbol{a}_m(x)$线性无关。

定义 3.3.2 设 $\boldsymbol{a}_1(x), \boldsymbol{a}_2(x), \cdots, \boldsymbol{a}_m(x)$是 m 个定义在$[a, b]$上的连续函数向量

$$\boldsymbol{a}_i(x)=(a_{i1}(x), a_{i2}(x), \cdots, a_{in}(x)) \quad (i=1, 2, \cdots, m)$$

记

$$g_{ij}\overset{\Delta}{=}\int_a^b \boldsymbol{a}_i(x)\boldsymbol{a}_j^{\mathrm{T}}(x)\mathrm{d}x \tag{3-6}$$

$$(i, j=1, 2, \cdots, m)$$

以 g_{ij} 为元素的常数矩阵

$$G=(g_{ij})_{m\times m}$$

称为 $\boldsymbol{a}_1(x), \boldsymbol{a}_2(x),\cdots,\boldsymbol{a}_m(x)$的 Gram 矩阵，det G 称为 Gram 行列式。

定理 3.3.1 定义在$[a, b]$上的连续函数向量 $\boldsymbol{a}_1(x), \boldsymbol{a}_2(x), \cdots, \boldsymbol{a}_m(x)$ 线性无关的充要条件是它的 Gram 矩阵为满秩矩阵。

〔证明〕 设

$$k_1\boldsymbol{a}_1(x)+k_2\boldsymbol{a}_2(x)+\cdots+k_m\boldsymbol{a}_m(x)=\boldsymbol{0}$$

在上式两边右乘 $\boldsymbol{a}_i^{\mathrm{T}}(x)$以后，对 x 积分得

$$\int_a^b\left[\sum_{j=1}^m k_j\boldsymbol{a}_j(x)\right]\boldsymbol{a}_i^{\mathrm{T}}(x)\mathrm{d}x=0 \qquad (i=1,2,\cdots,m) \tag{3-7}$$

上式左边即为

$$\sum_{j=1}^m k_j\int_a^b \boldsymbol{a}_j(x)\boldsymbol{a}_i^{\mathrm{T}}(x)\mathrm{d}x=\sum_{j=1}^m k_j g_{ji} \qquad (i=1,2,\cdots,m)$$

命

$$\boldsymbol{u}=(k_1, k_2, \cdots, k_m)^{\mathrm{T}}$$

式(3-7)可以写成

$$G^T u=0 \tag{3-8}$$

若 G^T 是满秩的，则式(3-8)只有零解，此即当 G 满秩时 $k_1=k_2=\cdots=k_m=0$，所以 $a_1(x)$，$a_2(x)$，…，$a_m(x)$ 线性无关。

若 G^T 不是满秩的，则式(3-8)有非零解，此即当 G 降秩时有一组不全为零的 k_1，k_2，…，k_m 存在，使得方程组(3-7)成立。以 k_1，k_2，…，k_m 依次乘方程组(3-7)的第一个，第二个，…，第 m 个方程并相加得

$$\int_a^b \left[\sum_{j=1}^m k_j a_j(x)\right]\left[\sum_{i=1}^m k_i a_i^T(x)\right]\mathrm{d}x=0 \tag{3-9}$$

命

$$a(x) \triangleq \sum_{i=1}^m k_i a_i(x) \tag{3-10}$$

将式(3-10)代入式(3-9)得

$$\int_a^b a(x)a^T(x)\mathrm{d}x=0 \tag{3-11}$$

因为 $a(x)$ 是连续的，$a^T(x)a(x)$ 对于所有的 x 都是非负的，所以只有当

$$a(x)=\sum_{i=1}^m k_i a_i(x)=0 \qquad \forall x\in[a,\ b]$$

时式(3-11)才成立，因而 $a_1(x)$，$a_2(x)$，…，$a_m(x)$ 是线性相关的。 ▌

根据定义判断一个函数向量组的线性相关性需要在区间 $[a,\ b]$ 上的每一点考察(3-5)式是否成立，因此是相当困难的。而定理 3.3.2 则给出一个简单的判定方法，即计算一个常数矩阵是否满秩就可以了。

例 3.3.1 设 $a_1(x)=(0,\ x)$，$a_2(x)=(x,\ 0)$，则

$$g_{11}=(a_1(x),a_1(x))=\int_a^b x^2\mathrm{d}x=\frac{1}{3}(b^3-a^3)$$

$$g_{12}=g_{21}=0$$

$$g_{22}=\int_a^b x^2\mathrm{d}x=\frac{1}{3}(b^3-a^3)$$

于是 $a_1(x)$，$a_2(x)$ 的 Gram 矩阵为

$$G=\begin{bmatrix}\frac{1}{3}(b^3-a^3) & 0\\ 0 & \frac{1}{3}(b^3-a^3)\end{bmatrix}$$

而

$$\det G=\frac{1}{9}(b^3-a^3)^2$$

故当 $a\neq b$ 时，$\det\ G>0$，$a_1(x)$，$a_2(x)$ 在 $[a,\ b]$ 上是线性无关的。

定理 3.3.1 指出，当函数向量组满足连续条件时，可以用 Gram 矩阵判断该函数向量组

是否线性无关。但 Gram 矩阵的元素是通过积分得到的，如果函数向量不仅连续，而且具有足够多阶的导数，那么可以通过对某个函数矩阵求导的方法来判断函数向量组的线性相关性。

定义 3.3.3 设 $\boldsymbol{a}_i(x)=(a_{i1}(x), a_{i2}(x), \cdots, a_{in}(x))(i=1, 2, \cdots, m)$ 是 m 个定义在$[a, b]$上的有 $m-1$ 阶导数的函数向量，记

$$A(x)\triangleq\begin{pmatrix}\boldsymbol{a}_1(x)\\ \boldsymbol{a}_2(x)\\ \vdots\\ \boldsymbol{a}_m(x)\end{pmatrix}=\begin{pmatrix}a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x)\\ a_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x)\\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots\\ a_{m1}(x) & a_{m2}(x) & \cdots & a_{mn}(x)\end{pmatrix} \tag{3-12}$$

那么称矩阵

$$W(x)\triangleq(A(x)\ \ A'(x)\cdots A^{(m-1)}(x))_{m\times mn}$$

$$=\begin{pmatrix}a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x) & \cdots\cdots & a_{11}^{(m-1)}(x) & a_{12}^{(m-1)}(x) & \cdots & a_{1n}^{(m-1)}(x)\\ a_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x) & \cdots\cdots & a_{21}^{(m-1)}(x) & a_{22}^{(m-1)}(x) & \cdots & a_{2n}^{(m-1)}(x)\\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots\\ a_{m1}(x) & a_{m2}(x) & \cdots & a_{mn}(x) & \cdots\cdots & a_{m1}^{(m-1)}(x) & a_{m2}^{(m-1)}(x) & \cdots & a_{mn}^{(m-1)}(x)\end{pmatrix}$$

是 $\boldsymbol{a}_1(x), \boldsymbol{a}_2(x), \cdots, \boldsymbol{a}_m(x)$的 Wronski 矩阵。其中 $A'(x)$、$A''(x)$、…、$A^{(m-1)}(x)$分别是 $A(x)$一阶、二阶、…、$(m-1)$阶导数矩阵。

下面的定理给出了函数向量组线性无关的一个充分条件。

定理 3.3.2 设 $W(x)$是 $\boldsymbol{a}_1(x), \boldsymbol{a}_2(x), \cdots, \boldsymbol{a}_m(x)$的朗斯基矩阵。若在区间$[a,b]$上的某个点$x_0\in[a, b]$，常数矩阵$W(x_0)$的秩等于 m。那么向量 $\boldsymbol{a}_1(x), \boldsymbol{a}_2(x), \cdots, \boldsymbol{a}_m(x)$在$[a, b]$上线性无关。

〔证明〕 用反证法证明。设 $\boldsymbol{a}_1(x), \boldsymbol{a}_2(x)\cdots, \boldsymbol{a}_m(x)$在$[a, b]$上线性相关，则存在不全为零的常数 $k_1, k_2, \cdots, k_m$ 使得对于任意$x\in[a, b]$都有

$$k_1\boldsymbol{a}_1(x)+k_2\boldsymbol{a}_2(x)+\cdots+k_m\boldsymbol{a}_m(x)=0$$

将此式逐次求导数，得

$$\begin{cases}k_1\boldsymbol{a}_1'(x)+k_2\boldsymbol{a}_2'(x)+\cdots+k_m\boldsymbol{a}_m'(x)=0\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ k_1\boldsymbol{a}_1^{(m-1)}(x)+k_2\boldsymbol{a}_2^{(m-1)}(x)+\cdots+k_m\boldsymbol{a}_m^{(m-1)}(x)=0\end{cases} \tag{3-13}$$

这 m 个等式对于任意$x\in[a, b]$均成立。以$x=x_0$代入(3-13)式，并改写成矩阵形式

$$\begin{pmatrix}\boldsymbol{a}_1(x_0) & \boldsymbol{a}_2(x_0) & \cdots & \boldsymbol{a}_m(x_0)\\ \boldsymbol{a}_1'(x_0) & \boldsymbol{a}_2'(x_0) & \cdots & \boldsymbol{a}_m'(x_0)\\ \cdots\cdots\cdots\cdots & \cdots\cdots & \cdots & \cdots\\ \boldsymbol{a}_1^{(m-1)}(x_0) & \boldsymbol{a}_2^{(m-1)}(x_0) & \cdots & \boldsymbol{a}_m^{(m-1)}(x_0)\end{pmatrix}\begin{pmatrix}k_1\\ k_2\\ \vdots\\ k_m\end{pmatrix}=0 \tag{3-14}$$

k_i 不全为零的充要条件是式(3-14)的系数矩阵的秩小于 m，此即 $\boldsymbol{a}_1(x)$，$\boldsymbol{a}_2(x)$，…，$\boldsymbol{a}_m(x)$ 的朗斯基矩阵 $W(x)$ 在 $x=x_0$ 处 $W(x_0)$ 的秩小于 m，与假设矛盾。 ∎

例 3.3.2 设

$$\boldsymbol{a}_1(x)=(1,\ x,\ x^2),\qquad \boldsymbol{a}_2(x)=(\mathrm{e}^x,\ 1,\ x)$$

则

$$A(x)=\begin{pmatrix}\boldsymbol{a}_1(x)\\ \boldsymbol{a}_2(x)\end{pmatrix}=\begin{pmatrix}1 & x & x^2\\ \mathrm{e}^x & 1 & x\end{pmatrix}$$

$$A'(x)=\begin{pmatrix}\boldsymbol{a}_1'(x)\\ \boldsymbol{a}_2'(x)\end{pmatrix}=\begin{pmatrix}0 & 1 & 2x\\ \mathrm{e}^x & 0 & 1\end{pmatrix}$$

$$W(x)=\begin{pmatrix}1 & x & x^2 & 0 & 1 & 2x\\ \mathrm{e}^x & 1 & x & \mathrm{e}^x & 0 & 1\end{pmatrix}$$

因为 $W(0)$ 的秩为2，所以 $\boldsymbol{a}_1(x)$ 与 $\boldsymbol{a}_2(x)$ 线性无关。

第四章　酉矩阵　Hermite矩阵

本章讨论在矩阵理论中占有重要地位的酉矩阵，Hermite矩阵和反Hermite矩阵。讨论的内容有：欧氏空间与酉空间，正交矩阵和酉矩阵，Schmidt正交化方法及其应用，二次齐式与对称矩阵，实对称矩阵，Hermite矩阵与Hermite齐式，正定Hermite矩阵，Rayleigh商。

§4.1　酉空间

1.　内积与酉空间

在线性空间中，向量之间的基本运算只有加法和数量乘法，统称为线性运算。如果我们以几何空间的向量作为线性空间的一个具体模型，那么向量的度量性质，如长度、夹角等在线性空间的理论中就没有得到反映。在这一节中，我们把度量的概念推广到一般的线性空间中。

在向量代数中，向量的度量性质都和向量的点积（数量积）紧密相关。而且向量的点积有明显的代数性质。所以在抽象的线性空间讨论中，我们取内积（以代替向量代数中的点积）作为基本的概念。

定义 4.1.1　设 V 是实数域 R 上的 n 维线性空间，V 中的任意两个向量 $\boldsymbol{x}$、$\boldsymbol{y}$ 用一个确定的方法使之对应于一个相应的实数，这个实数称为内积，记作 $(\boldsymbol{x},\ \boldsymbol{y})$，并且内积满足下列四个条件：

（1）　$(\boldsymbol{x},\ \boldsymbol{y})=(\boldsymbol{y},\ \boldsymbol{x})$

（2）　$(k\boldsymbol{x},\ \boldsymbol{y})=k(\boldsymbol{x},\ \boldsymbol{y})$，$k$ 为任何实数。

（3）　$(\boldsymbol{x}+\boldsymbol{y},\ \boldsymbol{z})=(\boldsymbol{x},\ \boldsymbol{z})+(\boldsymbol{y},\ \boldsymbol{z})$

（4）　$(\boldsymbol{x},\ \boldsymbol{x})\geqslant 0$，当且仅当 $\boldsymbol{x}=0$ 时 $(\boldsymbol{x},\ \boldsymbol{x})=0$。

这里 $\boldsymbol{x}$，$\boldsymbol{y}$ 和 $\boldsymbol{z}$ 是 V 中任意向量。这样的线性空间 V 称为 n 维欧几里得空间，简称 **n 维欧氏空间**。

例 4.1.1　在 n 维实向量空间 R^n 中对向量 $\boldsymbol{x}=(a_1,a_2,\cdots,a_n)^T$ 与 $\boldsymbol{y}=(b_1,b_2,\cdots,b_n)^T$ 定义内积为

$$(\boldsymbol{x},\ \boldsymbol{y})=\boldsymbol{y}^T\boldsymbol{x}=a_1b_1+a_2b_2+\cdots+a_nb_n \qquad (4\text{-}1)$$

不难验证，这样确定的内积符合定义 4.1.1 中的四个条件。所以 R^n 是欧氏空间，仍用 R^n 表示。公式（4-1）是欧氏空间 R^n 中最常用的内积定义。

例 4.1.2　在闭区间 $[a,\ b]$ 上所有连续函数构成的线性空间 $C_{[a,b]}$ 中，对函数 $f(t)$，$g(t)\in C_{[a,b]}$ 定义内积为

$$(f, g)=\int_a^b f(t)g(t)\mathrm{d}t$$

则 $C_{[a,b]}$ 是欧氏空间。

事实上，连续函数一定可积，所以 $\int_a^b f(t)g(t)\mathrm{d}t$ 是唯一确定的实数。同时它满足

（1）　$(f, g)=\int_a^b f(t)g(t)\mathrm{d}t=\int_a^b g(t)f(t)\mathrm{d}t=(g, f)$

（2）　$(kf,g)=\int_a^b kf(t)g(t)\mathrm{d}t=k\int_a^b f(t)g(t)\mathrm{d}t=k(f, g)$

（3）　$(f+g,h)=\int_a^b (f(t)+g(t))h(t)\mathrm{d}t$

$$=\int_a^b f(t)h(t)\mathrm{d}t+\int_a^b g(t)h(t)\mathrm{d}t$$

$$=(f, h)+(g, h)$$

（4）　$(f, f)=\int_a^b f(t)f(t)\mathrm{d}t=\int_a^b f^2(t)\mathrm{d}t\geqslant 0$，且 $\int_a^b f^2(t)\mathrm{d}t=0$ 的充要条件为 $f(t)\equiv 0$。因此 $C_{[a,b]}$ 是欧氏空间。

例 4.1.3　设在 R^2 中对向量 $\boldsymbol{x}=(a_1, a_2)^T$ 和 $\boldsymbol{y}=(b_1, b_2)^T$ 定义内积为

$$(\boldsymbol{x}, \boldsymbol{y})=2a_1b_1+a_1b_2+a_2b_1+a_2b_2$$

那么这样确定的内积符合定义 4.1.1 中的四个条件。

（1）　$(\boldsymbol{y}, \boldsymbol{x})=2b_1a_1+b_1a_2+b_2a_1+b_2a_2$

$$=2a_1b_1+a_1b_2+a_2b_1+a_2b_2=(\boldsymbol{x}, \boldsymbol{y})$$

（2）　$(k\boldsymbol{x}, \boldsymbol{y})=2ka_1b_1+ka_1b_2+ka_2b_1+ka_2b_2$

$$=k(2a_1b_1+a_1b_2+a_2b_1+a_2b_2)=k(\boldsymbol{x}, \boldsymbol{y})$$

（3）　设 $\boldsymbol{z}=(c_1, c_2)^T$，则

$$(\boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z})=2(a_1+b_1)c_1+(a_1+b_1)c_2+(a_2+b_2)c_1+(a_2+b_2)c_2$$

$$=2a_1c_1+a_1c_2+a_2c_1+a_2c_2+2b_1c_1+b_1c_2+b_2c_1+b_2c_2$$

$$=(\boldsymbol{x}, \boldsymbol{z})+(\boldsymbol{y}, \boldsymbol{z})$$

（4）　$(\boldsymbol{x}, \boldsymbol{x})=2a_1^2+a_1a_2+a_2a_1+a_2^2$

$$=(a_1+a_2)^2+a_1^2\geqslant 0$$

等式成立的充要条件是 $a_1=a_2=0$，即 $x=0$。

上面讨论的内积是在实数域上的线性空间内定义的，因此内积 (x, y) 是实数。对于复数域上的线性空间，我们规定两个向量 x，y 的内积 (x, y) 是一个复数，因此内积定义中的对称性条件 $(x, y)=(y, x)$ 要作相应改变。

定义 4.1.2 设 V 是复数域上的线性空间，V 中的任意两个向量 x，y 用一个确定的方法使之对应于一个相应的复数，这个复数称为内积，记作 (x, y)，并且内积满足下列四个条件：

（1） $(x, y)=\overline{(y, x)}$，其中 $\overline{(y, x)}$ 是 (y, x) 的共轭复数。

（2） $(x, ky)=k(x, y)$，k 为任意复数。

（3） $(x+y, z)=(x, z)+(y, z)$

（4） (x, x) 为非负实数，当且仅当 $x=0$ 时，$(x, x)=0$。

这里的 x、y、z 是 V 中任意向量，这样的线性空间称为**复欧氏空间**，简称**酉空间**。

显然酉空间是欧氏空间的推广。

例 4.1.4 在 n 维复向量空间 C^n 中对任意向量 $x=(a_1, a_2, \cdots, a_n)^T$ 与 $y=(b_1, b_2, \cdots, b_n)^T$ 定义内积为

$$(x, y)=(\overline{y})^T x=a_1\bar{b}_1+a_2\bar{b}_2+\cdots+a_n\bar{b}_n \tag{4-2}$$

不难验证，这样确定的内积符合定义 4.1.2 中的四条性质，所以 C^n 是酉空间，以后仍用 C^n 表示。公式（4-2）是酉空间 C^n 中最常用的内积定义。

2. 酉空间的性质

由于酉空间的性质与欧氏空间的性质很相似，有一套平行的理论。下面仅就酉空间情况讨论，其结论对欧氏空间基本适用，有细微不同之处请读者自行考虑。

由定义容易验证内积的性质：

（1） $(kx, y)=\bar{k}(x, y)$

（2） $(x, y+z)=(x, y)+(x, z)$

根据内积定义的第四个条件，对于酉空间和欧氏空间中的任何一个向量 x 都有 $(x, x)\geqslant 0$，因此利用 (x, x) 可以定义 x 的长度 $|x|$。

定义 4.1.3 称实数 $\sqrt{(x,x)}$ 为向量 x 的长度（或称向量 x 的模），记为 $|x|$，即

$$|x|=\sqrt{(x,x)} \tag{4-3}$$

向量 x 的长度 $|x|$ 具有通常的性质：$|kx|=|k||x|$，而且 $|x|=0$ 的充要条件是 $x=0$。

事实上

$$|kx|=\sqrt{(kx,kx)}=\sqrt{k\bar{k}(x,x)}=|k||x|$$

而且

$$|x|=0\Leftrightarrow(x, x)=0\Leftrightarrow x=0$$

根据向量长度的这个性质，我们可以将任何一个非零向量化成单位向量。因为对于任何

$x \neq 0$，总有

$$\left|\frac{x}{|x|}\right|=\frac{1}{|x|}|x|=1$$

也就是说，$\frac{x}{|x|}$是一个单位向量。用向量x的长度$|x|$除向量x，使之成为单位向量的过程称为向量x的单位化。

定义 4.1.4 若向量x，y的内积为零，即

$$(x,\ y)=0 \tag{4-4}$$

那么就称x，y正交或互相垂直，记之为$x \perp y$。

显然，只有零向量才和自己正交，也只有零向量能与任何一个向量正交。此外，当$x \perp y$时$(y,\ x)=0$也成立，即$x \perp y$与$y \perp x$是等价的。

由例 4.1.4 知，在R^n中$x \perp y$的充要条件是$y^T x=0$或$x^T y=0$。有时我们称R^n中两个正交向量是实正交的。

由例 4.1.4 知，在C^n中$x \perp y$的充要条件为$\bar{y}^T x=0$或$\bar{x}^T y=0$。有时，我们称C^n中两个正交向量是复正交的。

在欧氏空间和酉空间中同样有勾股定理，即当$x \perp y$时，

$$|x+y|^2=|x|^2+|y|^2 \tag{4-5}$$

事实上

$$\begin{aligned}|x+y|^2&=(x+y,\ x+y)\\&=(x,\ x)+(x,\ y)\\&\quad+(y,\ x)+(y,\ y)\\&=|x|^2+|y|^2\end{aligned}$$

用归纳法可把勾股定理推广到多个向量的情况，即当x_1，x_2，…，x_s两两正交时，总有

$$\begin{aligned}&|x_1+x_2+\cdots+x_s|^2\\&=|x_1|^2+|x_2|^2+\cdots+|x_s|^2\end{aligned}$$

定义 4.1.5 酉空间（或欧氏空间）中一组非零的向量，如果它们两两正交则称向量组为正交向量组。若正交向量组中的每一个向量还是单位向量，便称向量组为标准正交向量组。

一个正交向量组有如下重要特性。

定理 4.1.1 设x_1，x_2，…，x_s是n维酉空间（或欧氏空间）中正交向量组，那么x_1，x_2，…，x_s线性无关。

〔证明〕 设

$$k_1 x_1+k_2 x_2+\cdots+k_s x_s=0$$

以x_i（$i=1,\ 2,\ \cdots,\ s$）与上式作内积运算得到

$$k_1(x_i, x_1)+k_2(x_i, x_2)+\cdots+k_s(x_i, x_s)=0$$

因为 $(x_i, x_j)=\delta_{ij}|x_j|^2$（其中 δ_{ij} 是 Kronecker 记号，当 $i=j$ 时，$\delta_{ij}=1$；当 $i\neq j$ 时，$\delta_{ij}=0$）所以上式简化成

$$k_i(x_i, x_i)=0$$

从 $x_i\neq 0$ 便得

$$k_i=0 \quad (i=1, 2, \cdots, s)$$

所以 $x_1, x_2, \cdots, x_s$ 线性无关。]

这个定理说明，在 n 维空间（酉空间或欧氏空间）中，两两正交的非零向量不能超过 n 个。这个结论的几何意义是清楚的。例如，在平面上找不到三个两两垂直的非零向量，在几何空间中找不到四个两两垂直的非零向量。

定义 4.1.6 在 n 维空间(酉空间或欧氏空间)中，由 n 个正交向量组成的基称为**正交基**，若正交基的每一个向量是单位向量，便称为**标准正交基**。

设 V 是一个 n 维空间（酉空间或欧氏空间），在 V 中取一组基 $e_1, e_2, \cdots, e_n$，对于 V 中任意两个向量

$$x=a_1e_1+a_2e_2+\cdots+a_ne_n$$

$$y=b_1e_1+b_2e_2+\cdots+b_ne_n$$

由内积的性质得

$$\begin{aligned}(x, y)&=(a_1e_1+a_2e_2+\cdots+a_ne_n,\\&\qquad b_1e_1+b_2e_2+\cdots+b_ne_n)\\&=\sum_{i,j=1}^{n}\bar{a}_ib_j(e_i, e_j)\end{aligned} \tag{4-6}$$

命

$$g_{ij}=(e_i, e_j) \tag{4-7}$$

于是

$$g_{ij}=\bar{g}_{ji} \quad (i, j=1, 2, \cdots, n) \tag{4-8}$$

命

$$G=\begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1n}\\ g_{21} & g_{22} & \cdots & g_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ g_{n1} & g_{n2} & \cdots & g_{nn}\end{pmatrix} \tag{4-9}$$

在进一步讨论矩阵的特性之前，我们先来给出一个定义。

定义 4.1.7 设 A 为 $m\times n$ 阶复矩阵，用 $\bar{A}$ 表示以 A 的元素的共轭复数为元素组成的矩

阵，命

$$A^H=\bar{A}^T$$

则称 A^H 为**复共轭转置矩阵**，或 Hermite **共轭矩阵**。

不难验证有下列等式：

（1） $A^H=\overline{A^T}=\bar{A}^T$

（2） $(A+B)^H=A^H+B^H$

（3） $(kA)^H=\bar{k}A^H$

（4） $(AB)^H=B^HA^H$

（5） $(A^H)^H=A$

根据（4-8）式知道矩阵 G 满足

$$G^H=G \tag{4-10}$$

在这章第五节将研究形如（4－10）式的矩阵，这是一类十分重要的矩阵。

我们称形如（4-9）式的矩阵 G 为基 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，…，$\boldsymbol{e}_n$ 的**度量矩阵**。
设

$$\boldsymbol{x}=a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2+\cdots+a_n\boldsymbol{e}_n$$

$$=(\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n)\begin{pmatrix}a_1\\a_2\\\vdots\\a_n\end{pmatrix}$$

$$\boldsymbol{y}=b_1\boldsymbol{e}_1+b_2\boldsymbol{e}_2+\cdots+b_n\boldsymbol{e}_n$$

$$=(\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n)\begin{pmatrix}b_1\\b_2\\\vdots\\b_n\end{pmatrix}$$

若记

$$\alpha=(a_1,\ a_2,\ \cdots,\ a_n)^T$$

$$\beta=(b_1,\ b_2,\ \cdots,\ b_n)^T$$

则（4-6）式可改写成矩阵形式

$$(\boldsymbol{x},\ \boldsymbol{y})=\alpha^HG\beta \tag{4-11}$$

这表明向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积可由 $\boldsymbol{x}$，$\boldsymbol{y}$ 在某基上的坐标和该基的度量矩阵表示。

显然，两个不同基的度量矩阵是不同的，它们之间的关系由下述定理给出。

定理 4.1.2 设 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，…，$\boldsymbol{e}_n$ 和 $\boldsymbol{e}'_1$，$\boldsymbol{e}'_2$，…，$\boldsymbol{e}'_n$ 为 V 的两组基，A、B 分别为其

度量矩阵，基的过渡矩阵为 P
即

$$(e_1', e_2', \cdots, e_n')=(e_1, e_2, \cdots, e_n)P \tag{4-12}$$

则两个度量矩阵 A 与 B 满足

$$B=P^HAP \tag{4-13}$$

〔证明〕 将 n 阶矩阵 P 按列分块写为

$$P=(P_1, P_2, \cdots, P_n) \tag{4-14}$$

则由式（4-12）可知

$$e_i'=(e_1, e_2, \cdots, e_n)P_i \quad (i=1, 2, \cdots, n)$$

于是根据式（4-11）可得

$$(e_i', e_j')=P_i^HAP_j \tag{4-15}$$

这样基 e_1'，e_2'，…，e_n' 的度量矩阵 B 为

$$B=\begin{pmatrix}(e_1', e_1') & (e_1', e_2') & \cdots & (e_1', e_n')\\ (e_2', e_1') & (e_2', e_2') & \cdots & (e_2', e_n')\\ \cdots & \cdots & \cdots & \cdots\\ (e_n', e_1') & (e_n', e_2') & \cdots & (e_n', e_n')\end{pmatrix}$$

$$=\begin{pmatrix}P_1^HAP_1 & P_1^HAP_2 & \cdots & P_1^HAP_n\\ P_2^HAP_1 & P_2^HAP_2 & \cdots & P_2^HAP_n\\ \cdots & \cdots & \cdots & \cdots\\ P_n^HAP_1 & P_n^HAP_2 & \cdots & P_n^HAP_n\end{pmatrix}$$

$$=\begin{pmatrix}P_1^H\\ P_2^H\\ \vdots\\ P_n^H\end{pmatrix}A(P_1, P_2, \cdots, P_n)=P^HAP$$

】

定义 4.1.8 设 A、B 是 n 阶复矩阵，若存在一个满秩复矩阵 P，使得

$$B=P^HAP \tag{4-16}$$

我们便说 A 和 B 是**复相合的**。若 A、B 与 P 都是实矩阵，且有

$$B=P^TAP \tag{4-17}$$

便说 A 和 B 是**实相合的**。

显然欧氏空间两个基的度量矩阵为实相合，酉空间两个基的度量矩阵为复相合。

在解析几何中，向量 x，y 的夹角 $\langle x, y\rangle$ 的余弦可以通过数量积表示

$$\cos\langle x, y\rangle=\frac{(x,y)}{|x||y|} \tag{4-18}$$

为了在酉空间（或欧氏空间）中利用内积来引入夹角的概念，我们需要证明在空间中，式（4-18）右端值的绝对值不大于1。

定理 4.1.3 在酉空间（或欧氏空间）中，任何内积（x，y）满足不等式

$$|(x, y)|\leqslant|x||y| \quad \forall x, y\in V \tag{4-19}$$

当且仅当 x，y 线性相关时等号才成立。

不等式（4-19）就是著名的Cauchy-Schwarz 不等式。

〔证明〕 x 与 y 有一个为零向量时，结论是显然的。现设 x，y 都是非零向量。

设 $u=ax+by$，则

$$\begin{aligned}(u, u)&=(ax+by, ax+by)\\&=a\bar{a}(x, x)+\bar{a}b(x, y)+\bar{b}a(y, x)\\&\quad+\bar{b}b(y, y)\geqslant 0\end{aligned}$$

取 $a=(y, y)$，$b=-(y, x)$ 代入上式并消去正的因子（y，y）得

$$(y, y)(x, x)\geqslant(y, x)(x, y)=|(x, y)|^2$$

所以

$$|(x, y)|\leqslant|x||y|$$

等号成立的充要条件是 $u=0$，此即 x 与 y 线性相关。 ▌

例 4.1.5 R^n 中的 Cauchy-Schwarz 不等式为

$$|a_1b_1+a_2b_2+\cdots+a_nb_n|\leqslant\sqrt{a_1^2+\cdots+a_n^2}\ \sqrt{b_1^2+\cdots+b_n^2}$$

$C_{[a,b]}$ 中的 Cauchy-Schwarz 不等式为

$$\left|\int_a^b f(t)g(t)\mathrm{d}t\right|\leqslant\left(\int_a^b f^2(t)\mathrm{d}t\right)^{1/2}\left(\int_a^b g^2(t)\mathrm{d}t\right)^{1/2}$$

定义 4.1.9 非零向量 x、y 的夹角 $\langle x, y\rangle$ 规定为

$$\langle x, y\rangle=\arccos\frac{(x, y)}{|x||y|} \quad 0\leqslant\langle x, y\rangle\leqslant\pi$$

显然，当 x 与 y 正交时，$\langle x, y\rangle=\pi/2$。这与解析几何中的观念相一致。

根据 Cauchy-Schwarz 不等式，我们有三角形不等式：

$$|x+y|\leqslant|x|+|y|$$

事实上

$$
\begin{aligned}
|\boldsymbol{x}+\boldsymbol{y}|^2&=(\boldsymbol{x}+\boldsymbol{y},\ \boldsymbol{x}+\boldsymbol{y})\\
&=(\boldsymbol{x},\ \boldsymbol{x})+(\boldsymbol{x},\ \boldsymbol{y})+(\boldsymbol{y},\ \boldsymbol{x})+(\boldsymbol{y},\ \boldsymbol{y})\\
&\leqslant|\boldsymbol{x}|^2+|\boldsymbol{x}||\boldsymbol{y}|+|\boldsymbol{y}||\boldsymbol{x}|+|\boldsymbol{y}|^2\\
&=(|\boldsymbol{x}|+|\boldsymbol{y}|)^2
\end{aligned}
$$

所以

$$|\boldsymbol{x}+\boldsymbol{y}|\leqslant|\boldsymbol{x}|+|\boldsymbol{y}|$$

§4.2　正交矩阵与酉矩阵

1.　正交矩阵与酉矩阵的性质

在实矩阵中，正交矩阵是一类重要的矩阵。

定义 4.2.1　若 n 阶实矩阵 A 满足

$$A^TA=AA^T=I \tag{4-20}$$

则称 A 是**正交矩阵**。

将 n 阶正交矩阵 A 按列和行分块，写为

$$
A=(\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \cdots,\ \boldsymbol{a}_n)
=\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\\ \vdots\\ \boldsymbol{b}_n\end{pmatrix} \tag{4-21}
$$

把式（4-21）代入式（4-20）得

$$
\begin{pmatrix}\boldsymbol{a}_1^T\\ \boldsymbol{a}_2^T\\ \vdots\\ \boldsymbol{a}_n^T\end{pmatrix}(\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \cdots,\ \boldsymbol{a}_n)=\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\\ \vdots\\ \boldsymbol{b}_n\end{pmatrix}(\boldsymbol{b}_1^T,\ \boldsymbol{b}_2^T,\ \cdots,\ \boldsymbol{b}_n^T)=I_n
$$

即

$$
\begin{pmatrix}
\boldsymbol{a}_1^T\boldsymbol{a}_1 & \boldsymbol{a}_1^T\boldsymbol{a}_2 & \cdots & \boldsymbol{a}_1^T\boldsymbol{a}_n\\
\boldsymbol{a}_2^T\boldsymbol{a}_1 & \boldsymbol{a}_2^T\boldsymbol{a}_2 & \cdots & \boldsymbol{a}_2^T\boldsymbol{a}_n\\
\cdots & \cdots & \cdots & \cdots\\
\boldsymbol{a}_n^T\boldsymbol{a}_1 & \boldsymbol{a}_n^T\boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n^T\boldsymbol{a}_n
\end{pmatrix}
$$

$$
=\begin{pmatrix} \boldsymbol{b}_1\boldsymbol{b}_1^T & \boldsymbol{b}_1\boldsymbol{b}_2^T & \cdots & \boldsymbol{b}_1\boldsymbol{b}_n^T \\ \boldsymbol{b}_2\boldsymbol{b}_1^T & \boldsymbol{b}_2\boldsymbol{b}_2^T & \cdots & \boldsymbol{b}_2\boldsymbol{b}_n^T \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{b}_n\boldsymbol{b}_1^T & \boldsymbol{b}_n\boldsymbol{b}_2^T & \cdots & \boldsymbol{b}_n\boldsymbol{b}_n^T \end{pmatrix}
$$

$$
=\begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}
$$

于是我们有下面两组等式:

$$\boldsymbol{a}_i^T\boldsymbol{a}_j=\delta_{ij} \tag{4-22}$$

$$\boldsymbol{b}_i\boldsymbol{b}_j^T=\delta_{ij}, \quad (i, j=1, 2, \cdots, n) \tag{4-23}$$

式(4-22)表明正交矩阵的列向量组是标准正交向量组，式(4-23)表明正交矩阵的行向量组是标准正交向量组。这个性质称为**正交化条件**。反之，若实矩阵 A 的列向量 $\boldsymbol{a}_i$ 满足正交化条件，则有 $A^TA=I$，从而有 $A^TA=AA^T=I$，于是 A 为正交矩阵。同样地，若实矩阵 A 的行向量满足正交化条件，也可推出 A 是正交矩阵。于是我们得到如下的定理。

定理 4.2.1 设 A 为 n 阶实矩阵，那么 A 是正交矩阵的充要条件是 A的列（行）向量满足正交化条件。

在复矩阵中用 Hermite 共轭代替实矩阵的转置便有如下酉矩阵定义。

定义 4.2.2 若复矩阵 A 有

$$A^HA=AA^H=I \tag{4-24}$$

便称 A 为**酉矩阵**。若 A 是 n 阶酉矩阵，简记为 $A\in U^{n\times n}$。

显然实的酉矩阵便是正交矩阵，用 $A\in E^{n\times n}$ 表示 A 为 n 阶正交矩阵。

定理 4.2.2 设 A、B 均为 n 阶酉矩阵，则

(1) $A^{-1}=A^H\in U^{n\times n}$

(2) $|\det A|=1$

(3) $A^T\in U^{n\times n}$

(4) AB 与 BA 都是酉矩阵。

〔证明〕 (1) 由 $AA^H=A^HA=I$ 知 $A^{-1}=A^H$，又因 $A^H(A^H)^H=A^HA=I$，所以 $A^{-1}=A^H\in U^{n\times n}$。

(2) 由 $AA^H=I$ 得 $\det A\det A^H=1$，因此 $\det A\,\overline{\det A}=1$，所以 $|\det A|=1$。

(3) 因为 $A^T(A^T)^H=A^T\overline{A}=\overline{A^H}\overline{A}=\overline{A^HA}=I$，所以 $A^T\in U^{n\times n}$。

(4) 因为 $(AB)(AB)^H=ABB^HA^H=I$，所以 $AB\in U^{n\times n}$。同理 $BA\in U^{n\times n}$。 ▌

推论 4.2.1 设 A、B 均为 n 阶正交矩阵，则

(1) $A^{-1}=A^T\in E^{n\times n}$

(2) $\det A=\pm 1$

（3） $A^T\in E^{n\times n}$

（4） AB、BA都是正交矩阵。

〔证明〕 当A是正交矩阵时，$\det A$是实数，于是由$|\det A|=1$便得$\det A=\pm1$。（1），（3），与（4）请读者自己证明。 】

定理4.2.3 设A为n阶复矩阵，那么A是酉矩阵的充要条件是A的列向量满足正交化条件或A的行向量满足正交化条件。即矩阵

$$A=(\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \cdots,\ \boldsymbol{a}_n)=\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\\ \vdots\\ \boldsymbol{b}_n\end{pmatrix}$$

是酉矩阵的充要条件是

$$\boldsymbol{a}_i^H\boldsymbol{a}_j=\delta_{ij}$$

或

$$\boldsymbol{b}_i\boldsymbol{b}_j^H=\delta_{ij}$$

证明留给读者作为练习。

2. 酉矩阵的特征值

关于酉矩阵的特征值与特征向量有如下定理。

定理4.2.4 设A是n阶酉矩阵，则

（1） A的任何一个特征值之模（绝对值）等于1。

（2） A的相异特征值所对应的特征向量是互相正交的。

〔证明〕（1） 设λ为A的特征值，$\boldsymbol{x}$是λ对应的一个特征向量，则有

$$A\boldsymbol{x}=\lambda\boldsymbol{x}$$

所以

$$\boldsymbol{x}^HA^H=\bar{\lambda}\boldsymbol{x}^H$$

以$A\boldsymbol{x}$右乘上式两边便得

$$\boldsymbol{x}^HA^HA\boldsymbol{x}=\bar{\lambda}\boldsymbol{x}^HA\boldsymbol{x}$$

因为$\boldsymbol{x}$是A的特征向量且A是酉矩阵，所以

$$\boldsymbol{x}^H\boldsymbol{x}=\bar{\lambda}\lambda\boldsymbol{x}^H\boldsymbol{x}$$

由于$\boldsymbol{x}$是非零向量，故$\boldsymbol{x}^H\boldsymbol{x}\neq0$，因此

$$\bar{\lambda}\lambda=1$$

即酉矩阵A的特征值的模等于1。

（2） 设λ_i与λ_j为A的两个相异特征值，向量$\boldsymbol{x}_i$与$\boldsymbol{x}_j$满足

$$Ax_i=\lambda_i x_i, \qquad Ax_j=\lambda_j x_j$$

于是有

$$x_i^H A^H A x_j=\bar{\lambda}_i x_i^H \lambda_j x_j$$

即

$$x_i^H x_j=\bar{\lambda}_i \lambda_j x_i^H x_j$$

移项得

$$x_i^H x_j(1-\bar{\lambda}_i \lambda_j)=0$$

由于

$$(1-\bar{\lambda}_i \lambda_j)\lambda_i=\lambda_i-\bar{\lambda}_i \lambda_j \lambda_i=\lambda_i-\lambda_j \neq 0$$

因而

$$1-\bar{\lambda}_i \lambda_j \neq 0$$

于是

$$x_i^H x_j=0$$

这说明 λ_i 的特征向量与 λ_j 的特征向量正交。 ∎

对于正交矩阵的特征值与特征向量也有类似结论。

推论 4.2.2 设 A 是 n 阶正交矩阵，则

（1） A 的任何一个特征值之模（绝对值）等于 1 。

（2） A 的相异特征值所对应的特征向量是互相正交的。

证明留给读者作为练习。

3. 酉矩阵的标准形

下面我们来讨论酉矩阵的标准形，先证明一个引理，它在后面一些定理的证明中多次用到。

引理 4.2.1 设 $\alpha_1=(a_{11}, a_{21}, \cdots, a_{n1})^T$ 是模为 1 的 n 元复向量，那么存在酉矩阵 U，以 α_1 为它的第 i 个列向量。

〔证明〕 用构造法证明。考虑线性齐次方程

$$\alpha_1^H x=\bar{a}_{11}x_1+\bar{a}_{21}x_2+\cdots+\bar{a}_{n1}x_n=0$$

其中 $x=(x_1, x_2, \cdots, x_n)^T$。取方程任何一个模为 1 的解向量 α_2，显然 α_2 与 α_1 正交。

再考虑方程组

$$\alpha_1^H x=0, \quad \alpha_2^H x=0$$

取方程任何一个模为 1 的解向量 α_3，显然 α_3 分别与 α_1 和 α_2 正交。

继续这样下去，假定已经求得了 $n-1$ 个两两正交的单位向量 $\alpha_1, \alpha_2, \cdots, \alpha_{n-1}$。解方程组

$$a_1^H x=0,\ a_2^H x=0,\ \cdots,\ a_{n-1}^H x=0$$

取它的任何一个模为1的解向量 $\boldsymbol{a}_n$。于是所得的 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_{n-1}, \boldsymbol{a}_n$ 是标准正交向量组。那么 n 阶矩阵 $U=(\boldsymbol{a}_2, \boldsymbol{a}_3, \cdots, \boldsymbol{a}_{i-1}, \boldsymbol{a}_1, \boldsymbol{a}_{i+1}, \cdots, \boldsymbol{a}_n)$ 即为所求酉矩阵。]

比引理更一般的命题可参阅下一节定理 4.3.5。

现在可以讨论酉矩阵的标准形。

定理 4.2.5 设 A 是 n 阶酉矩阵，则一定存在 n 阶酉矩阵 U，使得

$$U^{-1}AU=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \tag{4-25}$$

〔证明〕 用数学归纳法证。A 的阶数为1时，$A=(e^{i\theta})$，命 $U=(e^{i\eta})$，则$U^{-1}=(e^{-i\eta})$，所以

$$U^{-1}AU=(e^{-i\eta})(e^{i\theta})(e^{i\eta})=(e^{i\theta})$$

设 A 的阶数为 $k-1$ 时定理成立，考虑阶数为 k 时的情况。

取 k 阶酉矩阵 A 的任何一个特征值 λ_1，对应的单位特征向量为 $\boldsymbol{a}_1$，根据引理 4.2.1，有一个以 $\boldsymbol{a}_1$ 为第一列的 k 阶酉矩阵 $U_1=(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k)$，因此

$$\begin{aligned}AU_1&=(A\boldsymbol{a}_1, A\boldsymbol{a}_2, \cdots, A\boldsymbol{a}_k)\\&=(\lambda_1\boldsymbol{a}_1, A\boldsymbol{a}_2, \cdots, A\boldsymbol{a}_k)\\&=(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k)\begin{pmatrix}\lambda_1 & a_2 & a_3 & \cdots & a_k\\ 0 & & & & \\ \vdots & & A_1 & & \\ 0 & & & & \end{pmatrix}\\&=U_1\begin{pmatrix}\lambda_1 & a_2 & a_3 & \cdots & a_k\\ 0 & & & & \\ \vdots & & A_1 & & \\ 0 & & & & \end{pmatrix}\end{aligned}$$

故得

$$U_1^{-1}AU_1=\begin{pmatrix}\lambda_1 & a_2 & a_3 & \cdots & a_k\\ 0 & & & & \\ \vdots & & A_1 & & \\ 0 & & & & \end{pmatrix} \tag{4-26}$$

根据定理 4.2.2 知 $U^{-1}A\,U$ 是酉矩阵，又由定理4.2.3 知它的每个行向量是单位向量，因而

$$\bar{\lambda}_1\lambda_1+\bar{a}_2a_2+\cdots+\bar{a}_ka_k=1$$

由定理 4.2.4 知

$$\bar{\lambda}_1\lambda_1=1$$

于是

$$\bar{a}_2a_2+\bar{a}_3a_3+\cdots+\bar{a}_ka_k=0$$

因为对于每一个 i 都有

$$\bar{a}_ia_i\geqslant 0 \quad (i=2,\ 3,\ \cdots,\ k)$$

所以

$$a_2=a_3=a_4=\cdots=a_k=0$$

代入式（4-26）得

$$U_1^{-1}AU_1=\begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & & & & \\ \vdots & & A_1 & & \\ 0 & & & & \end{pmatrix}$$

显然 A_1 是一个 $k-1$ 阶酉矩阵。由归纳假设，存在一个 $k-1$ 阶酉矩阵 W，使得

$$W^{-1}A_1W=\mathrm{diag}(\lambda_2,\ \lambda_3,\ \cdots,\ \lambda_k)$$

命

$$U_2=\begin{pmatrix} 1 & \\ & W \end{pmatrix}$$

则

$$U_2^{-1}U_1^{-1}AU_1U_2=\begin{pmatrix} 1 & \\ & W^{-1} \end{pmatrix}\begin{pmatrix} \lambda_1 & \\ & A_1 \end{pmatrix}\begin{pmatrix} 1 & \\ & W \end{pmatrix}$$

$$=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_k)$$

若命 $U=U_1U_2$，代入上式便得

$$U^{-1}AU=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_k)$$

显然 U 是酉矩阵。 ▮

若 A 是实的正交矩阵，根据定理 4.2.5 知，存在酉矩阵 U，使得 $U^{-1}AU=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)$，由矩阵相似对角形的理论知道，$\lambda_1$，$\lambda_2$，…，$\lambda_n$ 恰是 A 的 n 个特征值，U 的 n 个列向量恰是 λ_1，λ_2，…，λ_n 所对应的特征向量。当正交矩阵 A 的特征值全是实数时，A 的所有特征向量均为实向量，因此矩阵 U 实质上是实的正交矩阵。这就是下述推论。

推论 4.2.2 设 A 是 n 阶正交矩阵，则存在酉矩阵 U，使得

$$U^{-1}AU=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

当 A 的特征值全是实数时（由推论 4.2.1 知其值为±1），U 是正交矩阵。

在矩阵理论中，满足式（4-25）的矩阵 A 与对角形矩阵 $\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 称为酉相似。

定义 4.2.3 设 A，B 均为 n 阶复矩阵，若存在 n 阶酉矩阵 U，使得

$$U^{-1}AU=B$$

我们便说 A 与 B **酉相似**。若存在正交矩阵 Q，使得

$$Q^{-1}AQ=B$$

我们便说 A 与 B **正交相似**。

显然，推论 4.2.2 表明正交矩阵与对角形矩阵或酉相似，或正交相似。

§4.3 Schmidt 正交化方法

1. Schmidt 正交化方法

每一个 n 维线性空间的基由 n 个线性无关的向量构成，如果这个空间是酉空间（或欧氏空间），我们总是能构造一个标准正交基，这是酉空间（或欧氏空间）的基本定理。从一组线性无关的向量出发构造一组标准正交向量组的方法称为 Schmidt 方法。现介绍如下。

设 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，…，$\boldsymbol{a}_r$ 是 n 维酉空间（或欧氏空间）中 r 个线性无关的列向量，由这 r 个列向量生成的 r 维线性子空间记为 $L(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r)$，现求 r 维子空间 $L(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r)$ 中的一个标准正交基。分两步进行：

（1） 正交化

命

$$\begin{cases} \boldsymbol{b}_1=\boldsymbol{a}_1 \\ \boldsymbol{b}_2=\boldsymbol{a}_2-\dfrac{(\boldsymbol{b}_1,\boldsymbol{a}_2)}{(\boldsymbol{b}_1,\boldsymbol{b}_1)}\boldsymbol{b}_1 \\ \boldsymbol{b}_3=\boldsymbol{a}_3-\dfrac{(\boldsymbol{b}_1,\boldsymbol{a}_3)}{(\boldsymbol{b}_1,\boldsymbol{b}_1)}\boldsymbol{b}_1-\dfrac{(\boldsymbol{b}_2,\boldsymbol{a}_3)}{(\boldsymbol{b}_2,\boldsymbol{b}_2)}\boldsymbol{b}_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \boldsymbol{b}_r=\boldsymbol{a}_r-\dfrac{(\boldsymbol{b}_1,\boldsymbol{a}_r)}{(\boldsymbol{b}_1,\boldsymbol{b}_1)}\boldsymbol{b}_1-\cdots-\dfrac{(\boldsymbol{b}_{r-1},\boldsymbol{a}_r)}{(\boldsymbol{b}_{r-1},\boldsymbol{b}_{r-1})}\boldsymbol{b}_{r-1} \end{cases} \qquad (4\text{-}27)$$

容易验证

$$(\boldsymbol{b}_i,\boldsymbol{b}_j)=0 \quad (i\neq j)$$

（2） 单位化

命

$$c_1=\frac{b_1}{|b_1|},\quad c_2=\frac{b_2}{|b_2|},\quad \cdots,\quad c_r=\frac{b_r}{|b_r|} \tag{4-28}$$

显然，c_1，c_2，…，c_r是标准正交向量组。它构成子空间$L(a_1, a_2, \cdots, a_r)$的标准正交基。

定理4.3.1 从n维酉空间（或欧氏空间）中的任一组基a_1，a_2，…，a_n出发，可以通过一种确定的方法（称为Schmidt 正交化）构造出一组标准正交基。

例4.3.1 在R^4空间中，设

$$a_1=(1,\ -1,\ 1,\ -1),\ a_2=(5,\ 1,\ 1,\ 1),$$

$$a_3=(-3,\ -3,\ 1,\ -3)$$

求子空间$L(a_1, a_2, a_3)$的一个标准正交基。

〔解〕 应用Schmidt 正交化方法我们得到

$$b_1=a_1=(1,\ -1,\ 1,\ -1)$$

$$b_2=a_2-\frac{(b_1,a_2)}{(b_1,b_1)}b_1=a_2-b_1=(4,\ 2,\ 0,\ 2)$$

$$b_3=a_3-\frac{(b_1,a_3)}{(b_1,b_1)}b_1-\frac{(b_2,a_3)}{(b_2,b_2)}b_2$$

$$=a_3-b_1-b_2=(0,\ 0,\ 0,\ 0)$$

因为$b_3=0$，所以三个向量a_1，a_2，a_3必须是线性相关。显然a_1，a_2是线性无关的，所以$L(a_1, a_2, a_3)$是二维子空间，把b_1，b_2单位化，就可以得到这个子空间的一组标准正交基。

例4.3.2 在所有多项式构成的线性空间中，定义内积

$$(x,y)=\int_{-1}^{1}x(t)y(t)\mathrm{d}t$$

考虑无穷序列

$$x_1(t),\ x_2(t),\ \cdots,\ x_n(t),\ \cdots$$

其中$x_n(t)=t^n$，$(n=1,\ 2,\ \cdots)$。应用Schmidt 正交化方法可得正交向量（多项式）组为：

$$y_0(t)=x_0(t)=1$$

$$y_1(t)=x_1(t)-\frac{(y_0,x_1)}{(y_0,y_0)}y_0(t)=t$$

$$y_2(t)=x_2(t)-\frac{(y_0,x_2)}{(y_0,y_0)}y_0(t)-\frac{(y_1,x_2)}{(y_1,y_1)}y_1(t)=t^2-\frac{1}{3}$$

$$y_3(t)=\cdots=t^3-\frac{3}{5}t$$

$$y_4(t)=t^4-\frac{6}{7}t^2+\frac{3}{35}$$

$$y_5(t)=t^5-\frac{10}{9}t^3+\frac{5}{21}t$$

可以证明

$$y_n(t)=\frac{n!}{(2n)!}\frac{d^n}{dt^n}(t^2-1)^n \quad (n=1,\ 2,\ \cdots)$$

而 $y_0(t)$，$y_1(t)$，…的标准多项式是

$$\varphi_0(t)=\sqrt{\frac{1}{2}}$$

$$\varphi_1(t)=\sqrt{\frac{3}{2}}\,t$$

$$\varphi_2(t)=\frac{1}{2}\sqrt{\frac{5}{2}}(3t^2-1)$$

$$\varphi_3(t)=\frac{1}{2}\sqrt{\frac{7}{2}}(5t^3-3t)$$

$$\varphi_4(t)=\frac{1}{8}\sqrt{\frac{9}{2}}(35t^4-30t^2+3)$$

$$\varphi_5(t)=\frac{1}{8}\sqrt{\frac{11}{2}}(63t^5-70t^3+15t)$$

………………………………………………

在无穷维空间中称多项式

$$P_n(t)=\frac{(2n)!}{2^n(n!)^2}y_n(t)=\frac{1}{2^n n!}\frac{d^n}{dt^n}(t^2-1)^n$$

是 Legendre 多项式。

2. 矩阵的 UR 分解与 QR 分解

我们应用Schmidt 正交化方法来讨论关于矩阵的分解问题。

形如

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix}$$

其中 a_{ii} 均为正实数的矩阵称为正线上三角复矩阵。

形如

$$\begin{pmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ a_{n1} & \cdots & & a_{nn} \end{pmatrix}$$

其中 a_{ii} 均为正实数的矩阵称为**正线下三角复矩阵**。

定理 4.3.2 任何一个 n 阶满秩复矩阵 A 一定可以唯一地表示成一个正线上三角复矩阵 R 和一个酉矩阵 U 的乘积

$$A=UR \tag{4-29}$$

或者唯一地表示成一个正线下三角复矩阵 R_1 和一个酉矩阵 U_1 的乘积

$$A=R_1U_1 \tag{4-30}$$

〔证明〕 把矩阵 A 按列分块：

$$A=(\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \cdots,\ \boldsymbol{a}_n) \tag{4-31}$$

显然，$\boldsymbol{a}_1$，$\boldsymbol{a}_2$，…，$\boldsymbol{a}_n$ 是线性无关的。从 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，…，$\boldsymbol{a}_n$ 出发，应用Schmidt 方法求出标准正交向量组 $\boldsymbol{c}_1$，$\boldsymbol{c}_2$，…，$\boldsymbol{c}_n$，它们之间有关系：

$$\begin{cases} \boldsymbol{a}_1=k_{11}\boldsymbol{c}_1 \\ \boldsymbol{a}_2=k_{21}\boldsymbol{c}_1+k_{22}\boldsymbol{c}_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \boldsymbol{a}_n=k_{n1}\boldsymbol{c}_1+k_{n2}\boldsymbol{c}_2+\cdots+k_{nn}\boldsymbol{c}_n \end{cases} \tag{4-32}$$

其中

$$k_{rr}=|\boldsymbol{b}_r|>0 \quad (r=1,\ 2,\ \cdots,\ n)$$

把式（4-32）代入式（4-31）得到

$$\begin{aligned} A&=(\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \cdots,\ \boldsymbol{a}_n) \\ &=(k_{11}\boldsymbol{c}_1,\ k_{21}\boldsymbol{c}_1+k_{22}\boldsymbol{c}_2,\ \cdots, \\ &\quad k_{n1}\boldsymbol{c}_1+k_{n2}\boldsymbol{c}_2+\cdots+k_{nn}\boldsymbol{c}_n) \\ &=(\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_n)\begin{pmatrix} k_{11} & k_{21} & \cdots & k_{n1} \\ & k_{22} & \cdots & k_{n2} \\ & & \ddots & \vdots \\ & & & k_{nn} \end{pmatrix}=UR \end{aligned}$$

或

$$A=(\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \cdots,\ \boldsymbol{a}_n)$$

$$
=\begin{pmatrix} k_{11} & & & \\ k_{21} & k_{22} & & \\ \cdots & & \ddots & \\ k_{n1} & k_{n2} & \cdots & k_{nn} \end{pmatrix} \begin{pmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_n^T \end{pmatrix} = R_1 U_1
$$

因为 c_1，c_2，，…，c_n 是标准正交向量组，所以矩阵 $U=(c_1, c_2, \cdots, c_n)$ 与 $U_1=(c_1^T, c_2^T, \cdots, c_n^T)^T$ 都是酉矩阵。

现在证明唯一性。

设 A 有两种形如（4-29）的分解式：

$$
A=UR=\tilde{U}\tilde{R} \qquad (4\text{-}33)
$$

其中 U 和 $\tilde{U}$ 都是酉矩阵，R 和 $\tilde{R}$ 都是正线上三角复矩阵。由（4-33）式得

$$
R=U^{-1}\tilde{U}\tilde{R}=V\tilde{R} \qquad (4\text{-}34)
$$

式中 $V=U^{-1}\tilde{U}$ 是酉矩阵，我们证明 V 是单位矩阵，则唯一性得证。

设

$$
R=\begin{pmatrix} k_{11} & k_{21} & \cdots & k_{n1} \\ & k_{22} & \cdots & k_{n2} \\ & & \ddots & \\ & & & k_{nn} \end{pmatrix}
$$

$$
\tilde{R}=\begin{pmatrix} \tilde{k}_{11} & \tilde{k}_{21} & \cdots & \tilde{k}_{n1} \\ & \tilde{k}_{22} & \cdots & \tilde{k}_{n2} \\ & & \ddots & \\ & & & \tilde{k}_{nn} \end{pmatrix}
$$

$$
V=\begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ u_{n1} & u_{n2} & \cdots & u_{nn} \end{pmatrix}
$$

把 R、$\tilde{R}$ 与 V 代入式（4-34）。首先比较等号左右两边矩阵第一列的对应元素，得到

$$
k_{11}=u_{11}\tilde{k}_{11}, \quad 0=u_{i1}\tilde{k}_{11} \quad (i=2, 3, \cdots, n)
$$

因为 $\tilde{k}_{11}>0$，所以

$$
u_{21}=u_{31}=\cdots=u_{n1}=0
$$

但 V 是酉矩阵，它的列向量是单位向量，因此 $u_{11}=1$，故得

$$
V=\begin{pmatrix} 1 & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & u_{n2} & \cdots & u_{nn} \end{pmatrix}
$$

又由 V 的第一个列向量与其余的列向量是正交的得到

$$u_{12}=u_{13}=\cdots=u_{1n}=0$$

故得

$$V=\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & u_{22} & \cdots & u_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ 0 & u_{n2} & \cdots & u_{nn}\end{pmatrix}$$

其次，比较式（4-34）两边矩阵的第二列得到

$$u_{22}=1,\ u_{32}=u_{42}=\cdots=u_{n2}=0$$

$$u_{23}=u_{24}=\cdots=u_{2n}=0$$

继续比较式（4-34）两边矩阵的第三列，第四列，…，第 n 列，最后得到

$$V=I_n=U^{-1}\tilde{U}$$

代入式（4-34）就有

$$R=\tilde{R},\ U=\tilde{U}$$

这就证明了分解式（4-29）的唯一性。类似地可证分解式（4-30）的唯一性。 ▌

若矩阵 A 是满秩的实矩阵，从定理的证明过程中可见 U 与 U_1 实际上是正交矩阵，R 与 R_1 是实的三角矩阵，于是我们有下面的推论。

推论 4.3.1 任何一个 n 阶满秩实矩阵 A 一定可以唯一地表示成一个正线上三角实矩阵 R 和一个正交矩阵 Q 的乘积

$$A=QR$$

或者唯一地表示成一个正线下三角实矩阵 R_1 和一个正交矩阵 Q_1 的乘积

$$A=R_1Q_1$$

定理 4.3.2 称为复矩阵的 UR 分解，推论 4.3.1 称为实矩阵的 QR 分解，它们在矩阵计算中十分重要，在第五章将作更深入的介绍。

例 4.3.3 设

$$A=\begin{pmatrix}1 & 1 & 0\\ 1 & -1 & 1\\ 0 & 0 & 2\end{pmatrix}$$

写出 A 的 QR 分解。

〔解〕 命

$$\boldsymbol{\alpha}_1=(1,\ 1,\ 0)^T,\ \boldsymbol{\alpha}_2=(1,\ -1,\ 0)^T,\ \boldsymbol{\alpha}_3=(0,\ 1,\ 2)^T$$

由Schmidt正交化方法，得

$$b_1 = a_1 = (1,\ 1,\ 0)^T$$

$$b_2 = a_2 - \frac{(b_1, a_2)}{(b_1, b_1)} b_1 = a_2 = (1,\ -1,\ 0)^T$$

$$b_3 = a_3 - \frac{(b_1, a_3)}{(b_1, b_1)} b_1 - \frac{(b_2, a_3)}{(b_2, b_2)} b_2$$

$$= a_3 - \frac{1}{2} b_1 + \frac{1}{2} b_2 = (0,\ 0,\ 2)^T$$

单位化后，得

$$c_1 = \left(\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}},\ 0\right)^T = \frac{1}{\sqrt{2}} a_1$$

$$c_2 = \left(\frac{1}{\sqrt{2}},\ -\frac{1}{\sqrt{2}},\ 0\right)^T = \frac{1}{\sqrt{2}} a_2$$

$$c_3 = (0,\ 0,\ 1)^T = \frac{1}{2} a_3 - \frac{1}{4} a_1 + \frac{1}{4} a_2$$

由上面三个等式解出 a_1，a_2，a_3，得

$$a_1 = \sqrt{2}\, c_1$$

$$a_2 = \sqrt{2}\, c_2$$

$$a_3 = \frac{1}{\sqrt{2}} c_1 - \frac{1}{\sqrt{2}} c_2 + 2c_3$$

因此

$$A = (a_1,\ a_2,\ a_3)$$

$$= (c_1,\ c_2,\ c_3) \begin{pmatrix} \sqrt{2} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & 0 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & 0 & 2 \end{pmatrix} = QR$$

或

$$A = (a_1,\ a_2,\ a_3)$$

$$=\begin{pmatrix}\sqrt{2} & 0 & 0\\ 0 & \sqrt{2} & 0\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 2\end{pmatrix}\begin{pmatrix}c_1^T\\ c_2^T\\ c_3^T\end{pmatrix}$$

$$=\begin{pmatrix}\sqrt{2} & 0 & 0\\ 0 & \sqrt{2} & 0\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 2\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\\ 0 & 0 & 1\end{pmatrix}=R_1Q_1$$

定理 4.3.3 （**Schur 定理**） 任何一个 n 阶复矩阵 A 酉相似于一个上（下）三角矩阵，三角矩阵的对角线元素是 A 的特征值。

〔证明〕 用数学归纳法证。阶数为 1 时定理显然成立。现设 A 的阶数为 $k-1$ 时定理成立，考虑 A 的阶数为 k 时的情况。

取 k 阶矩阵 A 的一个特征值 λ_1，对应的单位特征向量为 $\boldsymbol{a}_1$，根据引理 4.2.1 有一个以 $\boldsymbol{a}_1$ 为第一列的 k 阶酉矩阵 $U_1=(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k)$，使得

$$AU_1=(A\boldsymbol{a}_1, A\boldsymbol{a}_2, \cdots, A\boldsymbol{a}_k)$$

$$=(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k)\begin{pmatrix}\lambda_1 & a_2 & \cdots & a_k\\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

故得

$$U^{-1}A\,U=\begin{pmatrix}\lambda_1 & a_2 & \cdots & a_k\\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

根据 纳假设 $k-1$ 阶矩阵 A_1 酉相似于上三角矩阵 R_1，即存在 $k-1$ 阶酉矩阵 W，使得

$$W^{-1}A\,W=R_1=\text{上三角矩阵}$$

命

$$U_2=\begin{pmatrix}1 & \\ & W\end{pmatrix}$$

则

$$U_2^{-1}U_1^{-1}AU_1U_2=\begin{pmatrix}\lambda_1 & b_2 & \cdots & b_k\\ 0 & & & \\ \vdots & & R_1 & \\ 0 & & & \end{pmatrix}=R$$

命 $U=U_1U_2$，它是酉矩阵，R 是上三角矩阵。因此

$$U^{-1}AU=R=\text{上三角矩阵}$$

因为 R 与 A 相似，R 是三角矩阵，所以三角矩阵 R 的对角线元素是 A 的 n 个特征值。 ▌

定理 4.3.4 在 n 维酉空间（或欧氏空间）中任取 m 个两两正交的单位向量 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$，则存在 $n-m$ 个单位向量 $\boldsymbol{\alpha}_{m+1}$，$\boldsymbol{\alpha}_{m+2}$，$\cdots$，$\boldsymbol{\alpha}_n$，使得 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\cdots$，$\boldsymbol{\alpha}_n$ 构成一组标准正交基。

〔证明〕 因为 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\cdots$，$\boldsymbol{\alpha}_m$ 线性无关、所以存在 $n-m$ 个向量 $\mathbf{b}_{m+1},\mathbf{b}_{m+2},\cdots,\mathbf{b}_n$，使得 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\cdots$，$\boldsymbol{\alpha}_m$，$\mathbf{b}_{m+1}$，$\mathbf{b}_{m+2}\cdots$，$\mathbf{b}_n$ 构成一组基。利用 Schmidt 正交化所得的前 m 个向量仍为 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\cdots$，$\boldsymbol{\alpha}_m$，再依次作下去，便能作出标准正交基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\cdots$，$\boldsymbol{\alpha}_m$，$\boldsymbol{\alpha}_{m+1}$，$\cdots$，$\boldsymbol{\alpha}_n$。 ▌

这个定理用矩阵语言叙述是：

定理 4.3.5 任给一个 $n\times m$（$m<n$）复矩阵

$$U_1=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1m}\\ a_{21} & a_{22} & \cdots & a_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ a_{n1} & a_{n2} & \cdots & a_{nm}\end{pmatrix}$$

它适合条件 $U_1^HU_1=I_m$，则必存在一个 $n\times(n-m)$ 矩阵

$$U_2=\begin{pmatrix}a_{1,m+1} & \cdots & a_{1n}\\ a_{2,m+1} & \cdots & a_{2n}\\ \cdots & \cdots & \cdots\\ a_{n,m+1} & \cdots & a_{nn}\end{pmatrix}$$

使得 n 阶矩阵

$$U=(U_1\quad U_2)$$

是一个酉矩阵。

〔证明〕 设 U_1 的 m 个列向量为

$$\boldsymbol{\alpha}_1=(a_{11}\ a_{21}\ \cdots\ a_{n1})^T,$$

$$\boldsymbol{\alpha}_2=(a_{12}\ a_{22}\ \cdots\ a_{n2})^T,$$

$$\cdots\cdots\cdots\cdots,$$

$$\boldsymbol{\alpha}_m=(a_{1m}\ a_{2m}\ \cdots\ a_{nm})^T$$

作为 n 维酉空间中的向量根据 $U_1^H U_1 = I_m$ 可知，α_1，α_2，…，α_m 均为单位向量且两两正交。由定理4.3.4，存在 $n-m$ 个单位向量 α_{m+1}，α_{m+2}，…，α_n，使得 α_1，α_2，…，α_m，α_{m+1}，…，α_n 构成一组标准正交基。再以 α_{m+1}，α_{m+2}，…，α_n 的坐标为列向量构成 $n\times(n-m)$ 矩阵 U_2，那么

$$U=(U_1 \quad U_2)$$

的列向量构成一组标准正交基，所以 U 是酉矩阵。 ▮

类似地，关于正交矩阵，我们有下述定理。

定理4.3.6 任给一个 $n\times m$ ($m<n$) 矩阵

$$Q_1=\begin{pmatrix} a_{11} & \cdots & a_{1m} \\ a_{21} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}$$

它适合条件 $Q_1^T Q_1 = I_m$，则必存在一个 $n\times(n-m)$ 矩阵

$$Q_2=\begin{pmatrix} a_{1,m+1} & \cdots & a_{1n} \\ a_{2,m+1} & \cdots & a_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots \\ a_{n,m+1} & \cdots & a_{nn} \end{pmatrix}$$

使得 n 阶矩阵

$$Q=(Q_1 \quad Q_2)$$

是一个正交矩阵。

在上一节，讨论了酉矩阵 A 与一个对角形矩阵酉相似，即对于任何一个酉矩阵 A，存在一个酉矩阵 U，使得 $U^{-1}AU$ 是对角形矩阵。现应用Schmidt正交化方法给出求酉矩阵 U 的步骤：

（1） 求出 A 的所有的相异特征值 λ_1，λ_2，…，λ_r，其重数分别为 k_1，k_2，…，k_r。

（2） 对于每一个 λ_j 求出 k_j 个线性无关的特征向量 α_{j1}，α_{j2}，…，α_{jk_j}。

（3） 应用Schmidt正交化方法，从 α_{j1}，α_{j2}，…，α_{jk_j} 出发求得标准正交向量组 c_{j1}，c_{j2}，…，c_{jk_j}，显然它们是 λ_j 的特征向量。

（4） 因为 λ_1，λ_2，…，λ_r 互异，所以向量组 c_{11}，c_{12}，…，c_{1k}，…，c_{r1}，c_{r2}，…，c_{rk_r} 是标准正交向量组，且由 $k_1+k_2+\cdots+k_r=n$ 知矩阵

$$U=(c_{11},\ c_{12},\ \cdots,\ c_{1k},\ \cdots,\ c_{r1},\ c_{r2},\ \cdots,\ c_{rk_r})$$

为所求之酉矩阵。

§4.4 二次齐式与对称矩阵

设 F 是一个数域，一个系数在数域 F 中的变量 x_1, x_2, …, x_n 的二次多项式

$$f(x_1, x_2, \cdots, x_n)$$
$$=a_{11}x_1^2+2a_{12}x_1x_2+\cdots+2a_{1n}x_1x_n+\cdots+a_{nn}x_n^2 \tag{4-35}$$

称为数域 F 上的一个 n 元二次齐式，简称二次齐式。

如果命

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad \mathbf{X}=(x_1, x_2, \cdots, x_n)^T$$

其中 $a_{ij}=a_{ji}\in F$，那么式（4-35）可写成

$$f(\mathbf{X})=\mathbf{X}^TA\mathbf{X} \tag{4-36}$$

这样，一个 n 元二次齐式与一个 n 阶对称矩阵一一对应，对称矩阵 A 称为二次齐式的矩阵，A 的秩称为二次齐式的秩。

设满秩矩阵 $C\in F^{n\times n}$，作线性变换

$$\mathbf{X}=C\mathbf{y} \tag{4-37}$$

其中 $\mathbf{y}=(y_1, y_2, \cdots, y_n)^T$，则二次齐式（4-35）变成

$$f(\mathbf{X})=\mathbf{X}^TA\mathbf{X}=f(C\mathbf{y})=\mathbf{y}^TC^TA\ C^T\mathbf{y}=\mathbf{y}^TB\mathbf{y}$$

其中

$$B=C^TAC \tag{4-38}$$

因为

$$B^T=(C^TAC)^T=C^TAC=B \tag{4-39}$$

所以，二次齐式在满秩线性变换 $\mathbf{X}=C\mathbf{y}$ 下仍变为二次齐式，其对应的矩阵满足式（4-39）。

定义 4.4.1 设 F 是数域，A、$B\in F^{n\times n}$，若存在满秩矩阵 $C\in F^{n\times n}$，使得

$$B=C^TA\ C$$

则称 A 与 B **相合**。

由定义 4.4.1 知，当 F 为实数域时，矩阵相合即为矩阵实相合（见定义4.1.8）。

二次齐式中最简单的一种是只包含平方项的二次齐式：

$$\mu_1y_1^2+\mu_2y_2^2+\cdots+\mu_ny_n^2 \tag{4-40}$$

形如式（4-40）的二次齐式称为标准形。对于任何一个二次齐式（4-36）是否存在一个满秩的线性变换（4-37），使得对于变量 y 而言化成形如（4-40）的二次齐式，这就是所谓化二次齐式为标准形的问题。

若用矩阵的语言来说，对于一个对称矩阵 A，是否存在一个满秩矩阵 C，使 C^TAC 为对角形矩阵。简言之，一个对称矩阵是否与对角形矩阵相合。下面的定理对这个问题给了肯定的回答。

定理 4.4.1 设 A 是数域 F 上 n 阶对称方阵，则有满秩矩阵 $C\in F^{n\times n}$，使得

$$C^TA\ C=\mathrm{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_n) \tag{4-41}$$

其中 $\mu_i\in F$，$(i=1,\ 2,\ \cdots,\ n)$

〔证明〕 用数学归纳法。

$n=1$ 时定理显然成立。今设 $n=k-1$ 时定理成立，讨论 $n=k$ 的情况。

（1） 设 $a_{11}\neq 0$，这时 A 可写为

$$A=\begin{pmatrix} a_{11} & \alpha^T \\ \alpha & A_1 \end{pmatrix}$$

其中 α 为 $(k-1)\times 1$ 矩阵，A_1 为 $(k-1)$ 阶对称矩阵。

令

$$C_1=\begin{pmatrix} 1 & -\dfrac{\alpha^T}{a_{11}} \\ 0 & I_{k-1} \end{pmatrix}$$

则有

$$C_1^TAC_1=\begin{pmatrix} a_{11} & 0 \\ 0 & A_1-\dfrac{\alpha\alpha^T}{a_{11}} \end{pmatrix}=\begin{pmatrix} a_{11} & 0 \\ 0 & \bar{A} \end{pmatrix}$$

因为 A_1 为 $(k-1)$ 阶对称矩阵，所以 $\bar{A}$ 是 $(k-1)$ 阶对称矩阵。从而由归纳假设，存在 $(k-1)$ 阶满秩矩阵 C_2，使得

$$C_2^T\bar{A}C_2=\mathrm{diag}(\mu_2,\ \mu_3,\ \cdots,\ \mu_k)$$

若令

$$C=C_1\begin{pmatrix} 1 & 0 \\ 0 & C_2 \end{pmatrix}$$

则有

$$C^TA\ C=\mathrm{diag}(\mu_1,\mu_2,\cdots,\mu_k)$$

（2） 设 $a_{11}=0$，但是对于某个 i 有 $a_{ii}\neq 0$，这时只要把 A 的第一行和第 i 行互换，然后再把第二列与第 i 列互换，就归结成情况 1。用矩阵乘法的形式就是对 A 施行如下运算（称相合变换）

$$C_1^T A C_1 = P(1, i)^T A P(1, i)$$

其中

$$C_1 = P(1,i) = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & & & \\ \vdots & & \ddots & \vdots & & & \\ & & 1 & & & & \\ 1 & & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 1 & & \\ \vdots & & & \vdots & & \ddots & \\ 0 & \cdots & \cdots & \vdots & \cdots & \cdots & 1 \end{pmatrix} \cdots i\text{行}$$

（第 i 列）

即初等矩阵 $P(1, i)$ 是由单位矩阵互换第一行与第 i 行所得。显然

$$P(1, i)^T = P(1, i)$$

（3） $a_{11}=a_{22}=\cdots=a_{nn}=0$，但有一个 $a_{1j}\neq 0$（$j\neq 1$），作相合变换

$$P(2, j)^T A P(2, j)$$

可以把 a_{1j} 移到第一行第二列的位置。所以可以认为 $a_{12}\neq 0$，这时可令

$$C_1 = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

于是 $C_1^T A\ C_1$ 的左上角就是

$$\begin{pmatrix} 2a_{12} & 0 \\ 0 & -2a_{12} \end{pmatrix}$$

归结为情况 1。

（4） $a_{11}=a_{12}=\cdots=a_{1n}=0$，由对称性知 $a_{11}=a_{21}=\cdots=a_{n1}=0$，于是

$$A = \begin{pmatrix} 0 & 0 \\ 0 & A_1 \end{pmatrix}$$

A_1 是（$k-1$）阶对称矩阵，由归纳假设有（$k-1$）阶满秩矩阵 C_1，使得

$$C_1^T A_1 C_1 = \mathrm{diag}(\mu_2, \mu_3, \cdots, \mu_k)$$

令

$$C=\begin{pmatrix}1 & 0\\ 0 & C_1\end{pmatrix}$$

则有

$$C^TA\ C=\text{diag}(0,\ \mu_2,\cdots,\ \mu_n)$$ ▮

推论 4.4.1 设 A 是数域 F 上的 n 阶对称矩阵，则 A 的相合对角形（4-41）中非零 μ_i 的个数为 A 的秩。

〔证明〕 由式（4-41）知

$$C^TA\ C=\text{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_n)$$

故

$$\begin{aligned}秩\ A&=秩\ \text{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_n)\\&=\{\mu_1,\ \mu_2,\ \cdots,\ \mu_n\}中非零的个数。\end{aligned}$$ ▮

若对称矩阵 A 的秩为 r，则 A 的相合对角形矩阵为

$$C^TA\ C=\text{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_r,\ 0,\ \cdots,\ 0) \tag{4-42}$$

式（4-42）称为数域 F 上对称矩阵 A 的相合标准形。

下面我们分别就 F 为复数域和实数域给出对称矩阵相合标准形的表示式。

定理 4.4.2 设 A 是复数域 F 上的 n 阶对称矩阵，$\text{rank}\ A=r$，则存在 n 阶复矩阵 C，使得

$$C^TA\ C=\text{diag}(I_r,\ 0) \tag{4-43}$$

〔证明〕 由于 A 的秩为 r，所以有 n 阶满秩矩阵 C_1，使得

$$C_1^TA\ C_1=\text{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_r,\ 0,\ \cdots,\ 0)$$

$\mu_i\neq 0\ (i=1,\ 2,\ \cdots,\ r)$。命

$$C_2=\text{diag}\left(\frac{1}{\sqrt{\mu_1}},\ \frac{1}{\sqrt{\mu_2}},\ \cdots,\ \frac{1}{\sqrt{\mu_r}},\ 1,\ \cdots,\ 0\right)$$

则 $C=C_1C_2$ 满足（4-43）式。 ▮

式（4-43）称为复数域上对称矩阵 A 的相合标准形。

定理 4.4.3 设 A 为实数域 F 上的 n 阶对称矩阵，$\text{rank A}=r$，则存在 n 阶实矩阵 C，使

$$C^TA\ C=\text{diag}(I_p,\ I_{r-p},\ 0) \tag{4-44}$$

这个定理的证明和定理 4.4.2 类似，只是在 $\mu_i<0$ 时，$\sqrt{\mu_i}$ 是虚数。不妨设 μ_1，μ_2，$\cdots$，μ_p 均大于 0。μ_{p+1}，$\cdots$，μ_r 均小于 0。令

$$C_2=\mathrm{diag}\left(\frac{1}{\sqrt{\mu_1}},\ \cdots,\ \frac{1}{\sqrt{\mu_p}},\ \frac{1}{\sqrt{-\mu_{p+1}}},\ \cdots,\ \frac{1}{\sqrt{-\mu_r}}\right)$$

参照定理 4.4 2，$C=C_1C_2$ 即为所求。

式（4-44）称为实数域上对称矩阵 A 的相合标准形。

如前所述，n 元二次齐式与 n 阶对称矩阵一一对应，而定理 4.4.3 表明一个实数域上的二次齐式可以化为如下形式：

$$y_1^2+y_2^2+\cdots+y_p^2-y_{p+1}^2-\cdots-y_r^2 \qquad (4\text{-}45)$$

形如（4-45）的二次齐式称为规范形。

定理 4.4.4（惯性定理） 任意一个实二次齐式 $f(\boldsymbol{X})=\boldsymbol{X}^TA\boldsymbol{X}$，经过一个适当的满秩线性变换 $\boldsymbol{X}=C\boldsymbol{y}$ 化为规范形

$$f(\boldsymbol{X})=y_1^2+y_2^2+\cdots+y_p^2-y_{p+1}^2-\cdots-y_r^2$$

则正项个数 p 与负项个数 $r-p$ 由矩阵 A 唯一确定。

〔证明〕 设 $f(\boldsymbol{X})$ 经过另一个满秩线性变换

$$\boldsymbol{X}=B\boldsymbol{z}$$

也化成规范形

$$f(\boldsymbol{X})=z_1^2+z_2^2+\cdots+z_q^2-z_{q+1}^2-\cdots-z_r^2$$

用反证法证明 $p=q$。不妨设 $p>q$。

显然，$\boldsymbol{y}$ 与 $\boldsymbol{z}$ 之间有关系

$$\boldsymbol{z}=B^{-1}C\boldsymbol{y}=G\boldsymbol{y}\overset{\Delta}{=}(g_{ij})_{n\times n}\boldsymbol{y}$$

把它写成方程组形式：

$$\begin{cases} z_1=g_{11}y_1+g_{12}y_2+\cdots+g_{1n}y_n \\ z_2=g_{21}y_1+g_{22}y_2+\cdots+g_{2n}y_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ z_n=g_{n1}y_1+g_{n2}y_2+\cdots+g_{nn}y_n \end{cases} \qquad (4\text{-}46)$$

考虑齐次线性方程组

$$\begin{cases} g_{11}y_1+g_{12}y_2+\cdots+g_{1n}y_n=0 \\ g_{21}y_1+g_{22}y_2+\cdots+g_{2n}y_n=0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ g_{q1}y_1+g_{q2}y_2+\cdots+g_{qn}y_n=0 \\ y_{p+1}=0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n=0 \end{cases} \qquad (4\text{-}47)$$

该方程组含有 n 个未知量，$q+(n-p)=n-(p-q)$ 个方程，所以必有非零解。不妨设方程组的一个非零解为

$$(y_1, y_2, \cdots, y_p, y_{p+1}, \cdots, y_n)$$
$$=(k_1, k_2, \cdots, k_p, k_{p+1}, \cdots, k_n)$$

于是我们有

$$f(X)=k_1^2+k_2^2+\cdots+k_p^2>0$$

另一方面又有

$$f(X)=-z_{q+1}^2-z_{q+2}^2-\cdots-z_r^2\leqslant 0$$

此为矛盾，故必有 $p\leqslant q$。

同理可证 $p\geqslant q$，于是 $p=q$。 ▮

二次齐式的规范形中，正项个数 p 称为**正惯性指数**，负项个数 $r-p$ 称为**负惯性指数**，它们的差 $p-(r-p)=2p-r$ 称为**符号差**。

§4.5 Hermite 矩阵与 Hermite 齐式

1. Hermite 矩阵

实对称矩阵是实矩阵中一类十分重要的矩阵，它在力学、物理学、自动控制与工程技术中有很广泛的应用。复矩阵中的 Hermite 矩阵与实矩阵中的实对称矩阵在其性质和证明方法上都十分相似。实对称矩阵在矩阵代数中已有了解，因此本节只介绍 Hermite 矩阵，而把实对称矩阵作为实数范围的 Hermite 矩阵处理。

定义 4.5.1 设 $A\in C^{n\times n}$，若 $A^H=A$，就称 A 是 **Hermite 矩阵**。若 $A^H=-A$，就称 A 是**反 Hermite 矩阵**。

当 A 的元素全为实数时，$A^H=A$，所以实对称矩阵是 Hermite 矩阵的特例。

由式（4-10）知，酉空间的度量矩阵是 Hermite 矩阵，欧氏空间的度量矩阵是实对称矩阵。

显然，Hermite 矩阵对角线元素全为实数，反 Hermite 矩阵对角线元素全为零。

根据复共轭转置运算的性质，当 A 与 B 是 n 阶 Hermite 矩阵时，$A+B$ 也是 Hermite 矩阵，但 AB 不一定是 Hermite 矩阵。

2. Hermite 矩阵的特征值与特征向量

定理 4.5.1 设 A 为 n 阶 Hermite 矩阵，则

（1） A 的特征值全是实数。

（2） A 的相异特征值所对应的特征向量是互相正交的。

〔证明〕（1） 设 λ 是 A 的特征值，x 是与 λ 对应的一个特征向量，即

$$Ax=\lambda x \tag{4-48}$$

于是有

$$\boldsymbol{x}^H A^H = \bar{\lambda}\boldsymbol{x}^H$$

因为 $A^H = A$，故

$$\boldsymbol{x}^H A = \bar{\lambda}\boldsymbol{x}^H \tag{4-49}$$

将 $\boldsymbol{x}^H$ 左乘式（4-48）两端，减去 $\boldsymbol{x}$ 右乘式（4-49）两端，得到

$$0 = (\lambda - \bar{\lambda})\boldsymbol{x}^H\boldsymbol{x}$$

由于 $\boldsymbol{x} \neq 0$，所以 $\boldsymbol{x}\ \boldsymbol{x} \neq 0$，故得

$$\lambda = \bar{\lambda}$$

此即 λ 是实数。

（2） 设 λ_i 与 λ_j 是 A 的两个相异特征值，向量 $\boldsymbol{x}_i$ 与 $\boldsymbol{x}_j$ 满足

$$A\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i \tag{4-50}$$

$$A\boldsymbol{x}_j = \lambda_j\boldsymbol{x}_j \tag{4-51}$$

将 $\boldsymbol{x}_j^H$ 左乘式（4-50）两端，得

$$\boldsymbol{x}_j^H A\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_j^H\boldsymbol{x}_i \tag{4-52}$$

将式（4-51）复共轭转置之后以 $\boldsymbol{x}_i$ 右乘两端，得

$$\boldsymbol{x}_j^H A\boldsymbol{x}_i = \boldsymbol{x}_j^H A\boldsymbol{x}_i = \lambda_j\boldsymbol{x}_j^H\boldsymbol{x}_i \tag{4-53}$$

由式（4-52）与式（4-53），得

$$(\lambda_i - \lambda_j)\boldsymbol{x}_j^H\boldsymbol{x}_i = 0$$

因为 λ_i 与 λ_j 互异，所以

$$\boldsymbol{x}_j^H\boldsymbol{x}_i = 0$$

即 $\boldsymbol{x}_i$ 与 $\boldsymbol{x}_j$ 正交。]

推论 4.5.1 实对称矩阵的特征值全是实数，实对称矩阵相异特征值所对应的特征向量是正交的。

定理 4.5.2 Hermite矩阵与对角形矩阵相似。

〔证明〕 用反证法。若 A 不能与对角形矩阵相似，那么它一定与若当矩阵相似，所以由定理 2.6.3 所述，至少对于 A 的某一个特征值 λ_i，有使

$$(\lambda I - A)\boldsymbol{x} \neq 0 \tag{4-54}$$

$$(\lambda I - A)^2\boldsymbol{x} = 0 \tag{4-55}$$

同时成立的 $\boldsymbol{x}$ 存在。

当 A 为 Hermite 矩阵时，$\lambda I - A$ 也是 Hermite 矩阵，以 $\boldsymbol{x}^H$ 左乘式（4-55）两边得

$$\boldsymbol{x}^H(\lambda I - A)^2\boldsymbol{x} = 0$$

即

$$[(\lambda I-A)\boldsymbol{x}]^H[(\lambda I-A)\boldsymbol{x}]=0$$

这表明

$$(\lambda I-A)\boldsymbol{x}=0$$

这与式(4-54)矛盾。 ∎

因为当 A 与对角形矩阵相似时 A 必有 n 个线性无关的特征向量，故有如下推论。

推论 4.5.2 n 阶 Hermite 矩阵有 n 个线性无关的特征向量。

根据推论 4.5.2 知，对于 Hermite 矩阵的任何一个 r 重特征值必有相应的 r 个线性无关的特征向量。应用 Schmidt 正交化方法，由这 r 个向量出发得到由 r 个特征向量组成的标准正交向量组。再由定理 3.5.2 便能得到下面的结论。

定理 4.5.3 任何一个 n 阶 Hermite 矩阵，必能找到由 n 个特征向量组成的标准正交向量组。

定理 4.5.4 设 A 是 n 阶复矩阵，那么 A 是 Hermite 矩阵的充要条件是存在 n 阶酉矩阵 U，使得

$$U^H A U=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \tag{4-56}$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 都是实数。

〔证明〕 设 A 为 n 阶 Hermite 矩阵，$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 是 A 的 n 个两两正交的单位特征向量，即

$$A\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i, \quad \boldsymbol{x}_i^H\boldsymbol{x}_j=\delta_{ij} \quad (i, j=1, 2, \cdots, n) \tag{4-57}$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 A 的 n 个实特征值。于是

$$\begin{aligned}
&A(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)\\
&=(A\boldsymbol{x}_1, A\boldsymbol{x}_2, \cdots, A\boldsymbol{x}_n)\\
&=(\lambda_1\boldsymbol{x}_1, \lambda_2\boldsymbol{x}_2, \cdots, \lambda_n\boldsymbol{x}_n)\\
&=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)
\end{aligned}$$

若命

$$U=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$$

显然，U 是一个 n 阶酉矩阵，代入上式得

$$AU=U\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

或

$$U^H A U=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

反之，若复矩阵 A 与实对角形矩阵 $\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 酉相似，即

$$U^H A U = \Lambda$$

或

$$A = U \Lambda U^H$$

易证

$$A^H = A$$

即 A 是 Hermite 矩阵。 ∎

实对称矩阵的特征值全是实数，从而它的特征向量必是实向量（Hermite 矩阵的特征向量就无此结果），因此式（4-56）中的酉矩阵 U 变为正交矩阵。这就证明了下面的推论。

推论 4.5.3 设 A 是 n 阶实矩阵，那么 A 是实对称矩阵的充要条件是存在 n 阶正交矩阵 Q，使得

$$Q^T A Q = \mathrm{diag}\,(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n) \tag{4-58}$$

其中 λ_1，λ_2，$\cdots$，λ_n 是实值。

定理 4.4.1 与定理 4.4.4 都告诉我们对称矩阵与一个对角形矩阵相合。下面我们将证明一个 Hermite 矩阵与一个对角形矩阵复相合。

定理 4.5.5 设 A 是秩为 r 的 Hermite 矩阵，那么存在满秩的 n 阶复矩阵 C，使得

$$C^H A C = \mathrm{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_r,\ 0,\ \cdots,\ 0) \tag{4-59}$$

其中 μ_1，μ_2，$\cdots$，μ_r 均为实数。

〔证明〕 用数学归纳法。注意到 Hermite 矩阵 $A = (a_{ij})$ 的对角线元素 a_{11}，a_{22}，$\cdots$，a_{nn} 都是实数。

$n=1$ 时定理显然成立。今设 $n=k+1$ 时定理成立，证明 $n=k$ 时定理为真。

（1） 假设 A 的对角线元素 a_{11}，a_{22}，$\cdots$，a_{nn} 中有一个不为零，不失一般性可设 $a_{11} \neq 0$，否则对于某个 i 有 $a_{ii} \neq 0$ 时，$P^H(1,\ i) A P(1,\ i)$ 中左上角元素 $a_{ii} \neq 0$。因此总可把 A 写成

$$A = \begin{pmatrix} a_{11} & \alpha^H \\ \alpha & A_1 \end{pmatrix}$$

其中 α 为 $(k-1)\times 1$ 矩阵，A_1 为 $k-1$ 阶 Hermite 矩阵。

令

$$C_1 = \begin{pmatrix} 1 & -\dfrac{\alpha^H}{a_{11}} \\ 0 & I_{k-1} \end{pmatrix}$$

则有

$$C_1^H A C_1 = \begin{pmatrix} a_{11} & 0 \\ 0 & A_1 - \dfrac{\alpha\alpha^H}{a_{11}} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 \\ 0 & \bar{A} \end{pmatrix}$$

显然 $\tilde{A}$ 是 $k-1$ 阶 Hermite 矩阵，由归纳假设，存在 $k-1$ 阶满秩矩阵 C_2，使得

$$C_2^H\tilde{A}C_2=\text{diag}\,(\mu_2,\ \mu_3,\ \cdots,\ \mu_k)$$

若令

$$C=C_1\begin{pmatrix}1 & 0\\ 0 & C_2\end{pmatrix}$$

则有

$$C^HA\ C=\text{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_k)$$

由于 A 与 $\text{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_k)$ 的秩相同，所以

$$C^HA\ C=\text{diag}\,(\mu_1,\ \mu_2,\ \cdots,\ \mu_r,\ 0,\ \cdots,\ 0)$$

（2） 若所有 $a_{ii}=0$，但 $a_{ij}\neq 0$，不失一般性，不妨设 $a_{12}\neq 0$。

命 $a=\bar{a}_{12}=a_{21}$，则 A 可以写成

$$A=\begin{pmatrix}P & B^H\\ B & A_1\end{pmatrix}$$

其中

$$P=\begin{pmatrix}0 & \bar{a}\\ a & 0\end{pmatrix}=P^H$$

B 为 $(k-2)\times 2$ 矩阵，A_1 为 $k-2$ 阶 Hermite 矩阵。若 $\text{Re}a\neq 0$，则令

$$C=\begin{pmatrix}Q & 0\\ 0 & I_{k-2}\end{pmatrix},\qquad Q=\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}=Q^H$$

便有

$$C^HA\ C=\begin{pmatrix}Q^H & 0\\ 0 & I_{k-2}\end{pmatrix}\begin{pmatrix}P & B^H\\ B & A_1\end{pmatrix}\begin{pmatrix}Q & 0\\ 0 & I_{k-2}\end{pmatrix}$$

$$=\begin{pmatrix}Q^HPQ & Q^HB^H\\ BQ & A_1\end{pmatrix}=\begin{pmatrix}Q^HPQ & (BQ)^H\\ BQ & A_1\end{pmatrix}$$

其中

$$Q^HP\ Q=\begin{pmatrix}a+\bar{a} & a-\bar{a}\\ \bar{a}-a & -(a+\bar{a})\end{pmatrix}$$

由于 $\text{Re}a\neq 0$，所以 $a+\bar{a}\neq 0$，这表明 Q^HPQ（或 $C^HA\ C$）的左上角元素不为零，归结为情况 1。

若 $\text{Re}a=0$，则令

$$a = \mathrm{i}\beta, \quad \beta \in R$$

在 k 阶矩阵 C 中的二阶子矩阵 Q 可改变为

$$Q = \begin{pmatrix} 1 & -\mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix} = Q^H$$

这时矩阵

$$Q^H P Q = \begin{pmatrix} 2\beta & 0 \\ 0 & -2\beta \end{pmatrix}$$

因此 $C^H A C$ 左上角元素不为零，又归结为情况 1。 】

3. Hermite 齐式

n 维复空间 C^n 内的 Hermite 齐式是指 n 个变量 $x_1, x_2, \cdots, x_n$ 和它们的共轭变量 $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n$ 之间的二次齐式

$$f(x_1, x_2, \cdots, x_n) = \sum_{i,j=1}^{n} a_{ij}\bar{x}_i x_j \qquad (4\text{-}60)$$

其中 $a_{ji} = \bar{a}_{ij}$，于是矩阵

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

具有性质 $A = A^H$，即 A 是 Hermite 矩阵。式（4-60）可写成矩阵形式

$$f(\boldsymbol{X}) = \boldsymbol{X}^H A \boldsymbol{X} \qquad (4\text{-}61)$$

其中 $\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$。$A$ 的秩称为Hermite 齐式的秩。显然，$f(\boldsymbol{X}) = \boldsymbol{X}^H A \boldsymbol{X}$ 总是实数。

我们知道，一个实对称矩阵与一个实二次齐式相对应。类似地，一个 Hermite 矩阵与一个 Hermite 二次齐式相对应。

若对 $\boldsymbol{X}$ 作可逆的的线性变换

$$\boldsymbol{X} = C\boldsymbol{y}, \quad \boldsymbol{y} = C^{-1}\boldsymbol{X} \qquad (4\text{-}62)$$

其中 C 为 n 阶满秩复矩阵，$\boldsymbol{y} = (y_1, y_2, \cdots, y_n)^{\mathrm{T}}$，则变换以后 $f(\boldsymbol{X})$ 成为

$$f(\boldsymbol{X}) = \boldsymbol{X}^H A \boldsymbol{X} = \boldsymbol{y}^H B \boldsymbol{y} \qquad (4\text{-}63)$$

其中

$$B = C^H A C \qquad (4\text{-}64)$$

且

$$B^H = B$$

所以对于变量 y 也是 Hermite 二次齐式。由式（4-64）知，变换前后的矩阵 A 与 B 满足复相合的关系。

Hermite 齐式中最简单的一种是只包含平方项的二次齐式

$$\mu_1 \bar{y}_1 y_1 + \mu_2 \bar{y}_2 y_2 + \cdots + \mu_n \bar{y}_n y_n \tag{4-65}$$

我们称形如式（4-65）的二次齐式是 Hermite 二次齐式的标准形。

定理 4.5.5 也可用二次齐式的语言来叙述。

定理 4.5.6 Hermite 二次齐式 $f(X) = X^H AX$ 总可以经过满秩的线性变换 $X = Cy$ 变二次齐式为平方和形式

$$\mu_1 \bar{y}_1 y_1 + \mu_2 \bar{y}_2 y_2 + \cdots + \mu_r \bar{y}_r y_r$$

其中 μ_1，μ_2，…，μ_r 都是实数，r 是 A 的秩。

Hermite 二次齐式与实二次齐式一样，有惯性定理。

定理 4.5.7 Hermite 二次齐式 $f(X) = X^H AX$ 总可以经过满秩的线性变换 $X = Cy$ 化为

$$f(X) = X^H AX = \bar{y}_1 y_1 + \bar{y}_2 y_2 + \cdots + \bar{y}_p y_p - \bar{y}_{p+1} y_{p+1} - \cdots - \bar{y}_r y_r \tag{4-66}$$

其中 r 为 A 的秩，p 由 A 所唯一确定。

定理的证明与定理 4.4.3 类似。式（4-66）称为 Hermite 二次齐式的规范形。称规范形中正项个数 p 为正惯性指数，负项个数 $r-p$ 为负惯性指数，它们的差 $p-(r-p)=2p-r$ 称为符号差。

§4.6 正定 Hermite 矩阵

前面已经指出，实二次齐式和 Hermite 二次齐式的标准形（或规范形）中正指数与负指数由二次齐式唯一确定。本节讨论正、负惯性指数与二次齐式值之间的关系。因为实二次齐式是 Hermite 齐式的特例，所以本节只讨论 Hermite 二次齐式。

因为正惯性指数 p 与秩 r 之间满足 $0 \leqslant p \leqslant r \leqslant n$，所以 Hermite 二次齐式可分为五种情况：（1）$p=0$，$r<n$；（2）$p=0$，$r=n$；（3）$0<p<r\ (\leqslant n)$；（4）$p=r<n$；（5）$p=r=n$。

现在分别讨论这五种情况：

（1） $p=0$，$r<n$。规范形为

$$X^H AX = -y_1 \bar{y}_1 - y_2 \bar{y}_2 - \cdots - y_r \bar{y}_r$$

不论 x_1，x_2，…，x_n 为何数，总有 $X^H AX \leqslant 0$，这时尽管 x_1，x_2，…，x_n 不全为零，仍有可能使 $X^H AX = 0$。例如，适合 $y_1 = y_2 = \cdots = y_r = 0$，而 y_{r+1}，y_{r+2}，…，y_n 不全为零的一组 x_1，x_2，…，x_n 的值就有 $X^H AX = 0$（这时 x_1，x_2，…，x_n 一定不全为零）。

（2） $p=0$，$r=n$。规范形为

$$X^HAX=-y_1\bar{y}_1-y_2\bar{y}_2-\cdots-y_n\bar{y}_n$$

只要 $y=(y_1, y_2, \cdots, y_n)$ 不全为零，就有 $X^HAX<0$，而只有在 $y=0$ 时，$X^HAX=0$。对于线性变换 $X=Cy$，因 $y=0$ 的充要条件是 $X=(x_1, x_2, \cdots, x_n)^T=0$，故二次齐式 X^HAX 之值（恒为实数）总是负数，只有 $X=0$ 时齐式之值才为零。

（3） $0<p<r\leqslant n$。规范形为

$$X^HAX=y_1\bar{y}_1+y_2\bar{y}_2+\cdots+y_p\bar{y}_p-y_{p+1}\bar{y}_{p+1}-\cdots-y_r\bar{y}_r$$

这时对于不同的 $X=(x_1, x_2, \cdots, x_n)^T$，二次齐式 X^HAX 之值可以大于零，小于零或等于零。

（4） $p=r<n$。规范形为

$$X^HAX=y_1\bar{y}_1+y_2\bar{y}_2+\cdots+y_r\bar{y}_r$$

这时，二次齐式之值 $X^HAX\geqslant 0$。

（5） $p=r=n$。规范形为

$$X^HAX=y_1\bar{y}_1+y_2\bar{y}_2+\cdots+y_n\bar{y}_n$$

这时，二次齐式 X^HAX 之值恒大于零。

根据上面的讨论，Hermite 二次齐式可分类如下。

定义 4.6.1 设有 Hermite 齐式（或实二次齐式）$f(X)=X^HAX$，则

（1） 对于任何 $X\neq 0$，$X^HAX>0$；$X^HAX=0$ 当且仅当 $X=0$，则称二次齐式 $f(X)=X^HAX$ 为正定的。

（2） 对于任何 $X\neq 0$，$X^HAX<0$；$X^HAX=0$ 当且仅当 $X=0$，则称二次齐式 $f(X)=X^HAX$ 为负定的。

（3） 若恒有 $X^HAX\geqslant 0$，且当 $X\neq 0$ 时也有 $X^HAX=0$，则称二次齐式 $f(X)=X^HAX$ 为半正定的

（4） 若恒有 $X^HAX\leqslant 0$，且当 $X\neq 0$ 时也有 $X^HAX=0$，则称二次齐式 $f(X)=X^HAX$ 为半负定的。

（5） 若 $f(X)=X^HAX$ 之值有时为正，有时为负，则称二次齐式 $f(X)=X^HAX$ 为不定的。

我们已经知道，一个 Hermite 齐式(实二次齐式)对应于唯一的 Hermite 矩阵（实对称矩阵）因此根据 Hermite 二次齐式（实二次齐式）之值的分类相应地对 Hermite 矩阵（实对称矩阵）进行分类。

定义 4.6.2 Hermite 矩阵(实对称矩阵)A 的正定、负定、半正定、半负定和不定，是由它所对应的二次齐式 X^HAX(或 X^TAX)的正定、负定、半正定、半负定和不定来定义的。

根据前面的讨论和上面的定义，容易推出下面的定理。

定理 4.6.1 对 Hermite 二次齐式（实二次齐式）X^HAX(X^TAX)，有

（1） 正定的充要条件为：$p=r=n$；

（2） 负定的充要条件为：$p=0$，$r=n$；

（3） 半正定的充要条件为：$p=r<n$；

（4） 半负定的充要条件为：$p=0$，$r<n$；

（5） 不定的充要条件为：$0<p<r(\leq n)$。

这里 p 是二次齐式 X^HAX 的正惯性指数，r 是秩。

定理 4.6.1 可也用矩阵语言来叙述。

定理 4.6.2 设 A 为 Hermite 矩阵（实对称矩阵）则

（1） A 为正定的充要条件为：存在满秩复（实）矩阵 C，使得

$$C^HA\ C=I_n$$

（或 $C^TA\ C=I_n$）

（2） A 为负定的充要条件为：存在满秩复（实）矩阵 C，使得

$$C^HA\ C=-I_n$$

（或 $C^TA\ C=-I_n$）。

（3） A 为半正定的充要条件为：存在满秩复（实）矩阵 C，使得

$$C^HA\ C=\mathrm{diag}\ (I_r,\ 0)$$

（或 $C^TA\ C=\mathrm{diag}(I_r,\ 0)$）。

（4） A 为半负定的充要条件为：存在满秩复（实）矩阵 C，使得

$$C^HA\ C=\mathrm{diag}\ (-I_r,\ 0)$$

（或 $C^TA\ C=\mathrm{diag}(-I_r,\ 0)$）。

（5） A 为不定的充要条件为：存在满秩复（实）矩阵 C，使得

$$C^HA\ C=\mathrm{diag}(I_p,\ -I_{r-p},\ 0)$$

（或 $C^TA\ C=\mathrm{diag}(I_p,\ -I_{r-p},\ 0)$）。

定理 4.6.3 设 A 为正定 Hermite（实对称）矩阵，那么 $\det A>0$。

〔证明〕 由定理 3.6.4 知，存在 C 使得

$$C^HA\ C=I_n$$

所以

$$\det C^H\ \det A\ \det C=1$$

故得

$$\det A=\frac{1}{\det C^H\ \det C}=\frac{1}{(\det C)^H(\det C)}>0$$

定理 4.6.4 设 A 是 n 阶 Hermite（实对称）矩阵，那么

（1） A 为正定的充要条件是 A 的特征值全是正实数。

（2） A 为半正定的充要条件是 A 的特征值全是非负实数。

（3） A 为负定的充要条件是 A 的特征值全是负实数。

（4） A 为半负定的充要条件是 A 的特征值全是非正实数。

〔证明〕 由定理 4.5.4 知存在酉（实正交）矩阵 U，使得

$$U^HA\ U=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)$$

其中 λ_1, λ_2, …, λ_n 是 A 的 n 个特征值，均是实数。也就是说，对于 Hermite（实）二次齐式 X^HAX 在线性变换 $X=Uy$ 下，简化成

$$X^HAX=\lambda_1y_1\bar{y}_1+\lambda_2y_2\bar{y}_2+\cdots+\lambda_ny_n\bar{y}_n$$

现证明 (1)。由正定定义知，对于任何 $X\neq0$，$X^HAX>0$。所以选 X 使得 $y_2=y_3=\cdots=y_n=0$，则仍有 $X^HAX=\lambda_1y_1\bar{y}_1>0$，因此 $\lambda_1>0$。类似地可得 $\lambda_2>0$，$\lambda_3>0$，…，$\lambda_n>0$。反之，若对于所有的 $\lambda_i>0$（$i=1, 2, \cdots, n$)，且 $y_i\bar{y}_i>0$，则对于任何的 X 都有 $X^HAX>0$。

(2)－(4) 的证明请读者完成。 ▌

定理 4.6.5 n 阶 Hermite（实对称）矩阵 A 正定的充要条件是存在 n 阶满秩复（实）矩阵 Q，使得

$$A=QQ^H$$

〔证明〕 由定理 4.6.2 知，A 正定的充要条件是存在满秩复（实）矩阵 C，使得

$$C^HA\ C=I_n$$

即

$$A=(C^H)^{-1}C^{-1}=(C^{-1})^HC^{-1}$$

命 $Q=(C^{-1})^H$ 代入上式得

$$A=QQ^H$$ ▌

对于半正定的矩阵也有类似的分解定理。

定理 4.6.6 n 阶 Hermite（实对称）矩阵 A 半正定的充要条件是存在 n 阶复（实）矩阵 Q，使得

$$A=QQ^H$$

矩阵 Q 的秩等于 A 的秩。

〔证明〕 A 半正定的充要条件是存在满秩复（实）矩阵 C，使得

$$C^HA\ C=\mathrm{diag}(I_r,\ 0)$$

其中 r 是 A 的秩，由上式有

$$\begin{aligned}A&=(C^H)^{-1}\mathrm{diag}(I_r,\ 0)C^{-1}\\&=(C^H)^{-1}\mathrm{diag}(I_r,\ 0)\mathrm{diag}(I_r,\ 0)C^{-1}\\&=[(C^H)^{-1}\mathrm{diag}(I_r,\ 0)][(C^H)^{-1}\mathrm{diag}(I_r,\ 0)]^H\end{aligned}$$

命 $Q=(C^H)^{-1}\mathrm{diag}(I_r,\ 0)$ 代入上式得

$$A=QQ^H$$

因为 Q 的秩等于 $\mathrm{diag}(I_r,\ 0)$ 的秩，所以 Q 的秩等于 A 的秩。 ▌

关于正定矩阵还有如下三角分解。

定理 4.6.7 设 A 是 n 阶 Hermite（实对称）矩阵，则 A 是正定矩阵的充要条件是存在正线上三角矩阵 R，使得

$$A=R^H R$$

或者存在正线下三角矩阵 L，使得

$$A=LL^H$$

〔证明〕 由定理 4.6.5 知存在满秩矩阵 Q，使得

$$A=QQ^H$$

由定理 4.3.2 知存在正线上三角矩阵 R 与酉矩阵 Q，使得

$$Q^H=UR$$

因此

$$A=R^H U^H U R=R^H R$$

反之，若 $A=R^H R$，由定理 4.6.5 便知 A 正定。

类似地可证 $A=LL^H$。]

定理 4.6.8 设 A 是 n 阶半正定（正定）Hermite 矩阵，则存在唯一的半正定（正定）Hermite 矩阵 H，使得

$$H^2=A$$

且任一和矩阵 A 可交换的复矩阵必和 H 可交换。

〔证明〕 存在酉矩阵 U，使得

$$A=U^H\text{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)U \tag{4-67}$$

其中 $0\leqslant\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$。所以令

$$H=U^H\text{diag}(\sqrt{\lambda_1},\ \sqrt{\lambda_2},\ \cdots,\ \sqrt{\lambda_n})U$$

显然 H 也是半正定 Hermite 矩阵。且 $H^2=A$。

现在证明唯一性。假设还有一个半正定 Hermite 矩阵 H_1 适合 $H_1^2=A$。对于 H_1，存在酉矩阵 U_1，使得

$$H_1=U_1^H\text{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_n)U_1$$

其中 $0\leqslant\mu_1\leqslant\mu_2\leqslant\cdots\leqslant\mu_n$，由此可知

$$A=H_1^2=U_1^H\text{diag}(\mu_1^2,\ \mu_2^2,\ \cdots,\ \mu_n^2)U_1$$

因此 $0\leqslant\mu_1^2\leqslant\mu_2^2\leqslant\cdots\leqslant\mu_n^2$ 是矩阵 A 的特征值，所以 $\mu_1^2=\lambda_1$，$\mu_2^2=\lambda_2$，$\cdots$，$\mu_n^2=\lambda_n$，即 $\mu_j=\sqrt{\lambda_j}$（$j=1,\ 2,\ \cdots,\ n$），故

$$H_1=U_1^H\text{diag}(\sqrt{\lambda_1},\ \sqrt{\lambda_2},\ \cdots,\ \sqrt{\lambda_n})U_1$$

由于 $A=H^2=H_1^2$，所以

$$A=U^H\mathrm{diag}\ (\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)U$$
$$=U_1^H\mathrm{diag}\ (\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)U_1$$

即

$$\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)UU_1^H$$
$$=UU_1^H\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)$$

命

$$UU_1^H=\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

代入上式得

$$\begin{pmatrix} \lambda_1 p_{11} & \lambda_1 p_{12} & \cdots & \lambda_1 p_{1n} \\ \lambda_2 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_2 p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_n p_{n1} & \lambda_n p_{n2} & \cdots & \lambda_n p_{nn} \end{pmatrix}=\begin{pmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{pmatrix}$$

即

$$\lambda_i p_{ij}=\lambda_j p_{ij}\ (i,\ j=1,\ 2,\ 3,\ \cdots,\ n)$$

当 $\lambda_i \neq \lambda_j$ 时，$p_{ij}=0$，所以

$$\sqrt{\lambda_i}\, p_{ij}=\sqrt{\lambda_j}\, p_{ij}\quad (i,\ j=1,\ 2,\ \cdots,\ n)$$

当 $\lambda_i=\lambda_j$ 时，等式

$$\sqrt{\lambda_i}\, p_{ij}=\sqrt{\lambda_j}\, p_{ij}\quad (i,\ j=1,\ 2,\ \cdots,\ n)$$

也成立。因此总有

$$\mathrm{diag}(\sqrt{\lambda_1},\ \sqrt{\lambda_2},\ \cdots,\ \sqrt{\lambda_n})UU_1^H$$
$$=UU_1^H\mathrm{diag}(\sqrt{\lambda_1},\ \sqrt{\lambda_2},\ \cdots,\ \sqrt{\lambda_n})$$

或

$$U^H\mathrm{diag}\ (\sqrt{\lambda_1},\ \sqrt{\lambda_2},\ \cdots,\ \sqrt{\lambda_n})U$$
$$=U_1^H\mathrm{diag}\ (\sqrt{\lambda_1},\ \sqrt{\lambda_2},\ \cdots,\ \sqrt{\lambda_n})U_1$$

此即

$$H=H_1$$

最后若矩阵 B 和 A 可交换，即 $AB=BA$，则

$$B(U^H\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)U)$$
$$=(U^H\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)U)B$$

所以

$$UBU^H\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$
$$=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)UBU^H$$

和证明唯一性相同可以得到

$$UBU^H\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n})$$
$$=\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n})UBU^H$$

即可得

$$BH=HB$$ ▮

注 1：不能用 $\sqrt{A}$ 来表示 H，而且，一般说来，如果不限定 H 为半正定（正定）Hermite 矩阵，那么矩阵方程 $A=H^2$ 的解 H 不唯一。

注 2：若 A 是正定，式（4-67）中的 $\lambda_i>0$，所以 H 正定，证明的其它地方完全相同。

若 A 是正定或半正定的实对称矩阵，也有类似结果。

定理 4.6.9 设 A 是 n 阶半正定（正定）实对称矩阵，则存在唯一的半正定（正定）实对称矩阵 H，使得

$$H^2=A,$$

且任一和矩阵 A 可交换的实矩阵必和 H 可交换。

证明与定理 4.6.7 几乎一字不差。

例 4.6.1 设两个 n 阶半正定的实对称矩阵

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad B=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix}$$

试证 n 阶矩阵

$$C=\begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}b_{n1} & a_{n2}b_{n2} & \cdots & a_{nn}b_{nn} \end{pmatrix}$$

也是半正定实对称矩阵。

〔证明〕 显然 C 是 n 阶实对称矩阵。任取 $\boldsymbol{X}=(x_1, x_2, \cdots, x_n)^T$，由于 A、B 半正定，所以

$$X^{T}AX=\sum_{i,j=1}^{n}a_{ij}x_i x_j\geqslant 0$$

$$X^{T}BX=\sum_{i,j=1}^{n}b_{ij}x_i x_j\geqslant 0$$

另外，由定理 4.6.6 知，存在 n 阶矩阵 $Q=(q_{ij})$，使得

$$B=QQ^{T}$$

即

$$b_{jk}=\sum_{l=1}^{n}q_{jl}q_{kl}\quad (j,\ k=1,\ 2,\ \cdots,\ n)$$

所以

$$\sum_{i,j=1}^{n}a_{ij}b_{ij}x_i x_j=\sum_{i,j=1}^{n}a_{ij}\left(\sum_{l=1}^{n}q_{il}q_{jl}\right)x_i x$$

$$=\sum_{l=1}^{n}\sum_{i,j=1}^{n}a_{ij}(x_i q_{il})(x_j q_{jl})$$

因为对于任何固定的 l，由 A 的半正定性知

$$\sum_{i,j=1}^{n}a_{ij}(x_i q_{il})(x_j q_{jl})\geqslant 0$$

所以

$$\sum_{i,j=1}^{n}a_{ij}b_{ij}x_i x_j\geqslant 0$$

此即 C 是半正定实对称矩阵。 ∎

为了证明正定矩阵的行列式法则，先引入正定矩阵的一个引理。

引理 4.6.1 设 A 为 n 阶正定 Hermite（实对称）矩阵，P 为 n 阶满秩矩阵，则 $P^{H}AP$ 也正定。

〔证明〕 存在满秩矩阵 Q，使得

$$A=QQ^{H}$$

所以

$$P^{H}A\,P=P^{H}Q\,Q^{H}P=(P^{H}Q)(P^{H}Q)^{H}$$

因为 $P^{H}Q$ 是满秩的。所以又由定理 4.6.5 知 $P^{H}AP$ 是正定矩阵。 ∎

若 A 为半正定的，那么 $P^{H}A\,P$ 也是半正定的。证明可仿照引理 4.6.1 的。

定义 4.6.3 设实对称矩阵

$$A=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix}$$

称如下的 n 个行列式

$$|a_{11}|,\quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},\quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},\quad \cdots,$$

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

为矩阵 A 的 n 个顺序主子式。

定理 4.6.10 n 阶 Hermite（实对称）矩阵 $A=(a_{ij})$ 正定的充要条件是 A 的 n 个顺序主子式都大于零。

〔证明〕 必要性 设 Hermite（实）二次齐式

$$f(\boldsymbol{X})=\boldsymbol{X}^H A\boldsymbol{X}=\sum_{i,j=1}^{n} a_{ij}\bar{x}_i x_j$$

是正定的。

对于每个 k，$1\leqslant k\leqslant n$，令

$$f_k(x_1,\ x_2,\ \cdots,\ x_k)=\sum_{i,j=1}^{k} a_{ij}\bar{x}_i x_j$$

我们来证 f_k 是一个 k 元的正定二次齐式。对于任意一组不全为零的复数 c_1，c_2，…，c_k 有

$$f_k(c_1,\ c_2,\ \cdots,\ c_k)=\sum_{i,j=1}^{k} a_{ij}\bar{c}_i c_j$$

$$=f(c_1,\ c_2,\ \cdots,\ c_k,\ 0,\ \cdots,\ 0)>0$$

因此 $f_k(x_1,\ x_2,\ \cdots,\ x_k)$ 是正定的。由定理 4.6.3 知 f_k 所对应的矩阵的行列式

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ \cdots & \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}>0 \qquad (k=1,\ 2,\ \cdots,\ n)$$

即 A 的 n 个顺序主子式都大于零。

充分性 对 A 的阶数 n 用归纳法。

$n=1$ 时结论显然成立。

设对于一切 $n-1$ 阶 Hermite 矩阵结论成立。下面证明阶数为 n 时结论也成立。

将矩阵 A 写成分块形式

$$A=\begin{pmatrix} A_1 & a \\ a^H & a_{nn} \end{pmatrix}$$

其中 A_1 为 $n-1$ 阶 Hermite 矩阵，a 为 $(n-1)\times 1$ 矩阵。根据假设 A 的顺序主子式全大于零，故 A_1 的顺序主子式也全大于零。根据归纳假设，A_1 是正定矩阵，因此存在满秩的 $n-1$ 阶矩阵 G，使得

$$G^H A_1 G = I_{n-1}$$

令

$$C_1 = \begin{pmatrix} G & 0 \\ 0 & 1 \end{pmatrix}$$

于是

$$C_1^H A C_1 = \begin{pmatrix} G^H & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 & a \\ a^H & a_{nn} \end{pmatrix} \begin{pmatrix} G & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} I_{n-1} & G^H a \\ a^H G & a_{nn} \end{pmatrix}$$

再令

$$C_2 = \begin{pmatrix} I_{n-1} & -G^H a \\ 0 & 1 \end{pmatrix}$$

则有

$$C_2^H C_1^H A C_1 C_2 = \begin{pmatrix} I_{n-1} & 0 \\ -a^H G & 1 \end{pmatrix} \begin{pmatrix} I_{n-1} & G^H a \\ a^H G & a_{nn} \end{pmatrix} \begin{pmatrix} I_{n-1} & -G^H a \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} I_{n-1} & 0 \\ 0 & a_{nn} - a^H G G^H a \end{pmatrix}$$

其中 $a_{nn} - a^H G\ G^H a = b =$ 实数。若令 $C = C_1 C_2$，则有

$$C^H A C = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & b \end{pmatrix}$$

两边取行列式

$$|C|^2 |A| = b$$

由假设 $|A|>0$，故 $b>0$。由于

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & b \end{pmatrix} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \sqrt{b} \end{pmatrix} \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \sqrt{b} \end{pmatrix}$$

因此 A 与 I_n 复相合。于是根据引理 4.6.1 知 A 是正定矩阵。 ∎

比定理 4.6.10 更一般的结果是如下定理。

定理 4.6.11 n 阶 Hermite（实对称）矩阵 $A=(a_{ij})$ 正定的充要条件是它的所有主子式[*]全大于零。

---

[*] 所谓主子式就是行列式 $|A|$ 中行指标与列指标相同的子式。

〔证明〕 由定理4.6.10知充分性显然成立。现证必要性。对于A的任一k阶子式

$$A\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ i_1 & i_2 & \cdots & i_k \end{pmatrix}=\begin{vmatrix} a_{i_1i_1} & a_{i_1i_2} & \cdots & a_{i_1i_k} \\ a_{i_2i_1} & a_{i_2i_2} & \cdots & a_{i_2i_k} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i_ki_1} & a_{i_ki_2} & \cdots & a_{i_ki_k} \end{vmatrix}$$

存在某个矩阵P（是k个初等变换矩阵$P(i, i_i)$的乘积。$i=1, 2, \cdots, k$），使得P^HAP的k阶顺序主子式为$A\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ i_1 & i_2 & \cdots & i_k \end{pmatrix}$，根据引理3.6.13知$P^HAP$正定，所以

$A\begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ i_1 & i_2 & \cdots & i_k \end{pmatrix}>0$。 ∎

§4.7 Rayleigh商

在这里我们只就复Hermite矩阵的特征值和Rayleigh商进行讨论，结论对实对称矩阵完全适用。

定义4.7.1 设A为n阶Hermite矩阵，又$\boldsymbol{X}$为n维复向量，称实数

$$R(\boldsymbol{X})=\frac{\boldsymbol{X}^HA\boldsymbol{X}}{\boldsymbol{X}^H\boldsymbol{X}} \quad (\boldsymbol{X}\neq\boldsymbol{0}) \tag{4-68}$$

为Hermite矩阵A的Rayleigh商。

Rayleigh商是研究特征值的一个十分有力的工具，例如研究Hermite二次齐式$\boldsymbol{X}^HA\boldsymbol{X}$在条件$\boldsymbol{X}^H\boldsymbol{X}=1$下的极值问题归结为研究$A$的特征值的极值问题。

设A为n阶Hermite矩阵，把A的特征值按递减顺序排列

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n \tag{4-69}$$

定理4.7.1 $R(\boldsymbol{X})$具有

（1） $R(k\boldsymbol{X})=R(\boldsymbol{X})$，$k$为任意实数。

（2） $\lambda_1\geqslant R(\boldsymbol{X})\geqslant\lambda_n$ （4-70）

（3） $\min\limits_{\boldsymbol{X}\neq\boldsymbol{0}} R(\boldsymbol{X})=\lambda_n$， $\max\limits_{\boldsymbol{X}\neq\boldsymbol{0}} R(\boldsymbol{X})=\lambda_1$

〔证明〕（1） 由定义得

$$R(k\boldsymbol{X})=\frac{(k\boldsymbol{X})^HA(k\boldsymbol{X})}{(k\boldsymbol{X})^H(k\boldsymbol{X})}=\frac{\bar{k}k\boldsymbol{X}^HA\boldsymbol{X}}{\bar{k}k\boldsymbol{X}^H\boldsymbol{X}}=R(\boldsymbol{X})$$

（2） 因为A是Hermite矩阵，所以存在酉矩阵U，使得

$$U^HAU=\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

命　　$X=Uy$

则　　$$R(X)=\frac{X^HAX}{X^HX}=\frac{y^HU^HAUy}{y^Hy}=\frac{y^H\Lambda y}{y^Hy}$$

所以对于任何 y 都有

$$\lambda_1\geqslant\frac{y^H\Lambda y}{y^Hy}=\frac{X^HAX}{X^HX}\geqslant\lambda_n$$

（3） 由（2）可见，$R(X)$ 能达到极值 λ_1、λ_n，即

$$\min_{X\neq 0} R(X)=\lambda_n,\quad \max_{X\neq 0} R(X)=\lambda_1$$

定理 4.7.1 告诉我们对于 Rayleigh 商 $R(X)$ 只要在单位球面 $X^HX=1$ 上考虑即可，且给出了 $R(X)$的最大值与最小值。

定理 4.7.2 设 $L(X_1, X_2, \cdots, X_{k-1})$ 表示由 A 的特征向量 $X_1, X_2, \cdots, X_{k-1}$ 生成的子空间，它的正交补子空间(*)用 R_k 表示，则

$$\lambda_k=\max_{X\in R_k} R(X) \tag{4-71}$$

其中 X_i 是 A 的对应于 λ_i 的特征向量。

〔证明〕 因为 A 是 Hermite 矩阵，不妨设 $X_1, X_2, \cdots, X_{k-1}, \cdots, X_n$ 为两两正交的单位特征向量组，对于任何向量X，都有

$$X=c_1X_1+c_2X_2+\cdots+c_nX_n$$

从而

$$\begin{aligned}R(X)&=\frac{(c_1X_1+\cdots+c_nX_n)^HA(c_1X_1+\cdots+c_nX_n)}{(c_1X_1+\cdots+c_nX_n)^H(c_1X_1+\cdots+c_nX_n)}\\&=\frac{(\bar c_1X_1^H+\cdots+\bar c_nX_n^H)(c_1\lambda_1X_1+\cdots+c_n\lambda_nX_n)}{\bar c_1c_1+\bar c_2c_2+\cdots+\bar c_nc_n}\\&=\frac{\bar c_1c_1\lambda_1+\bar c_2c_2\lambda_2+\cdots+\bar c_nc_n\lambda_n}{\bar c_1c_1+\bar c_2c_2+\cdots+\bar c_nc_n}\\&=a_1\lambda_1+a_2\lambda_2+\cdots+a_n\lambda_n\end{aligned}$$

其中　　$$a_i=\frac{\bar c_ic_i}{\sum_{j=1}^n\bar c_jc_j}\geqslant 0,\ 且\sum_{i=1}^n a_i=1 \tag{4-72}$$

当 $k=1$ 时，$R_1=c^n$。此即定理 4.7.1。

当 $k=2$ 时，$X\in R_2$，这时 $c_1=0$，

$$X=c_2X_2+\cdots+c_nX_n$$

(*) 设 V_1、V_2 是欧氏空间 V 的两个子空间，如果对于任意的 $a\in V_1$，$b\in V_2$，恒有 $(a, b)=0$，并且 $V_1+V_2=V$，便称 V_2 是 V_1 的正交补子空间。

从而 $$R(\boldsymbol{X})=a_2\lambda_2+\cdots+a_n\lambda_n$$

因为a_2，a_3，…，a_n的取值取决于$\boldsymbol{X}$，且满足式（4-72），所以

$$\max_{\boldsymbol{X}\in R_2} R(\boldsymbol{X})=\lambda_2$$

类似地推理便得式（4-71）。 ▮

定理 4.7.3 设$\boldsymbol{X}\in L(\boldsymbol{X}_r, \boldsymbol{X}_{r+1}, \cdots, \boldsymbol{X}_s)$，$1\leqslant r, s\leqslant n$，则

$$\max_{\boldsymbol{X}\neq 0} R(\boldsymbol{X})=\lambda_r, \quad \min_{\boldsymbol{X}\neq 0} R(\boldsymbol{X})=\lambda_s \tag{4-73}$$

〔证明〕 设

$$\boldsymbol{X}=c_r\boldsymbol{X}_r+c_{r+1}\boldsymbol{X}_{r+1}+\cdots+c_s\boldsymbol{X}_s$$

由定理 3.8.3 的证明可知

$$R(\boldsymbol{X})=a_r\lambda_r+\cdots+a_s\lambda_s$$

因此

$$\max_{\boldsymbol{X}\in L(\boldsymbol{X}_r,\cdots,\boldsymbol{X}_s)} R(\boldsymbol{X})=\lambda_r, \quad \min_{\boldsymbol{X}\in L(\boldsymbol{X}_r,\cdots,\boldsymbol{X}_s)} R(\boldsymbol{X})=\lambda_s$$ ▮

定理 4.7.4 设V_k是复空间中任意k维子空间，则有极大-极小原理

$$\lambda_k=\max_{V_k}\min_{\boldsymbol{X}\in V_k} R(\boldsymbol{X}) \tag{4-74}$$

或极小-极大原理

$$\lambda_k=\min_{V_{n-k+1}}\max_{\boldsymbol{X}\in V_{n-k+1}} R(\boldsymbol{X}) \tag{4-75}$$

〔证明〕 $k-1$维子空间$L(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{k-1})$的正交补子空间$R_k$是$n-k+1$维空间，因此任意$k$维子空间$V_k$与$R_k$必有公共的非零向量$\boldsymbol{y}_k$。由定理 4.7.2 知

$$\lambda_k=\max_{\boldsymbol{X}\in R_k} R(\boldsymbol{X})\geqslant R(\boldsymbol{y}_k)$$

又因$\boldsymbol{y}_k\in V_k$，所以

$$\min_{\boldsymbol{X}\in V_k} R(\boldsymbol{X})\leqslant R(\boldsymbol{y}_k)\leqslant\lambda_k$$

它对于任何k维子空间都成立，因此

$$\max_{V_k}\min_{\boldsymbol{X}\in V_k} R(\boldsymbol{X})\leqslant\lambda_k$$

另一方面，由定理 4.7.1 知

$$\lambda_k=\min_{\boldsymbol{X}\in L(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_k)} R(\boldsymbol{X})\leqslant\max_{V_k}\min_{\boldsymbol{X}\in V_k} R(\boldsymbol{X})$$

综合上述两不等式得到

$$\lambda_k = \max_{V_k} \min_{X \in V_k} R(X)$$

现在利用式（4-74）来证明式（4-75）。

令 $B=-A$，B 的特征值按递减顺序排列为

$$\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_n$$

其中 $\mu_k = -\lambda_{n-k+1}$，（$k=1, 2, \cdots, n$），对于矩阵 B 应用式（4-74）有

$$\lambda_{n-k+1} = -\mu_k = -\max_{V_k} \min_{0 \neq X \in V_k} \frac{X^H B X}{X^H X}$$

$$= -\max_{V_k} \left\{ \min_{0 \neq X \in V_k} \frac{-X^H A X}{X^H X} \right\}$$

$$= -\max_{V_k} \left\{ -\max_{0 \neq X \in V_k} \frac{X^H A X}{X^H X} \right\}$$

$$= \min_{V_k} \max_{0 \neq X \in V_k} \frac{X^H A X}{X^H X}$$

$$= \min_{V_k} \max_{X \in V_k} R(X)$$

将（$n-k+1$）以 i 代替便得

$$\lambda_i = \min_{V_{n-i+1}} \max_{X \in V_{n-i+1}} R(X)$$

例 4.7.1 设 A、B 都是正定 Hermite 矩阵，试证

$$\lambda_{\min}(A+B) \geqslant \lambda_{\min}(A) + \lambda_{\min}(B)$$

〔证明〕

$$\lambda_{\min}(A+B) = \min \frac{X^H (A+B) X}{X^H X}$$

$$= \min\left(\frac{X^H A X}{X^H X} + \frac{X^H B X}{X^H X} \right)$$

$$\geqslant \min \frac{X^H A X}{X^H X} + \min \frac{X^H B X}{X^H X}$$

$$= \lambda_{\min}(A) + \lambda_{\min}(B)$$

]

在这节最后简单介绍 Hermite 矩阵特征值的振动定理，即讨论矩阵的元素发生微小变化时对应矩阵特征值的变化范围。

定理 4.7.5 设 A、E 均为 n 阶 Hermite 矩阵，矩阵 $\tilde{A}=A+E$。设 λ_1，λ_2，$\cdots$，λ_n；μ_1，μ_2，$\cdots$，μ_n 与 $\tilde{\lambda}_1$，$\tilde{\lambda}_2$，$\cdots$，$\tilde{\lambda}_n$ 分别是矩阵 A、E 和 $\tilde{A}$ 的按递减顺序排列的特征

值，则

$$\lambda_i+\mu_n\leqslant\tilde{\lambda}_i\leqslant\lambda_i+\mu_1 \quad (i=1,2,\cdots,n) \tag{4-76}$$

〔证明〕 设酉矩阵 $U=(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)$ 满足

$$U^H A U=\operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

命

$$V_k=L(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_k)$$

由定理 4.7.4 得

$$\begin{aligned}
\tilde{\lambda}_k&=\max_{V_k}\min_{\boldsymbol{X}\in V_k}\frac{\boldsymbol{X}^H\tilde{A}\boldsymbol{X}}{\boldsymbol{X}^H\boldsymbol{X}}\geqslant\min_{\boldsymbol{X}\in V_k}\frac{\boldsymbol{X}^H\tilde{A}\boldsymbol{X}}{\boldsymbol{X}^H\boldsymbol{X}}\\
&=\min_{\boldsymbol{X}\in V_k}\left(\frac{\boldsymbol{X}^HA\boldsymbol{X}}{\boldsymbol{X}^H\boldsymbol{X}}+\frac{\boldsymbol{X}^HE\boldsymbol{X}}{\boldsymbol{X}^H\boldsymbol{X}}\right)\\
&\geqslant\min_{\boldsymbol{X}\in V_k}\frac{\boldsymbol{X}^HA\boldsymbol{X}}{\boldsymbol{X}^H\boldsymbol{X}}+\min_{\boldsymbol{X}\in V_k}\frac{\boldsymbol{X}^HE\boldsymbol{X}}{\boldsymbol{X}^H\boldsymbol{X}}\\
&=\lambda_k+\min_{\boldsymbol{X}\in V_k}\frac{\boldsymbol{X}^HE\boldsymbol{X}}{\boldsymbol{X}^H\boldsymbol{X}}\geqslant\lambda_k+\mu_n
\end{aligned} \tag{4-77}$$

命 $\quad A=\tilde{A}+(-E)$

显然 $\quad \lambda_{\min}(-E)=-\mu$

根据式（4-77）得

$$\lambda_k\geqslant\tilde{\lambda}_k-\mu_1 \tag{4-78}$$

综合式（4-77）与式（4-78）得

$$\lambda_k+\mu_n\leqslant\tilde{\lambda}_k\leqslant\lambda_k+\mu_1$$ ∎

习　　题

4-1 设 $A=(a_{ij})$ 是一个 n 阶正定 Hermite 矩阵，在 n 维复线性空间 C 中向量

$$\boldsymbol{x}=(x_1, x_2, \cdots, x_n), \quad \boldsymbol{y}=(y_1, y_2, \cdots, y_n)$$

定义内积（$\boldsymbol{x}, \boldsymbol{y}$）为

$$(\boldsymbol{x}, \boldsymbol{y})=\boldsymbol{x}A\boldsymbol{y}^T$$

（1） 证明在这个定义下，C 成为一个欧氏空间。

（2） 求单位向量 $\boldsymbol{e}_1=(1, 0, \cdots, 0), \cdots, \boldsymbol{e}_n=(0, 0, \cdots, 0, 1)$ 的度量矩阵。

（3） 具体写出这个空间中的 Cauchy-Schwarz 不等式。

4-2 设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 是 n 维欧氏空间 V 的一组基，证明

（1） 如果$\boldsymbol{a}\in V$使得$(\boldsymbol{a}, \boldsymbol{a}_i)=0$，$(i=1, 2, \cdots, n)$，则，$\boldsymbol{a}=\boldsymbol{0}$。

（2） 如果$\boldsymbol{a}, \boldsymbol{b}\in V$，且$(\boldsymbol{a}, \boldsymbol{a}_i)=(\boldsymbol{b}, \boldsymbol{a}_i)$，$(i=1, 2, \cdots, n)$，则$\boldsymbol{a}=\boldsymbol{b}$。

4-3 设$\boldsymbol{e}_1$，$\boldsymbol{e}_2$，$\boldsymbol{e}_3$，$\boldsymbol{e}_4$，$\boldsymbol{e}_5$是是五维欧氏空间V的一组标准正交基，$V_1=L(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3)$，求$V_1$的一组标准正交基。其中$\boldsymbol{a}_1=\boldsymbol{e}_1+\boldsymbol{e}_5$，$\boldsymbol{a}_2=\boldsymbol{e}_1-\boldsymbol{e}_2+\boldsymbol{e}_4$，$\boldsymbol{a}_3=2\boldsymbol{e}_1+\boldsymbol{e}_2+\boldsymbol{e}_3$。

4-4 求齐次线性方程组

$$\begin{cases}2x_1+x_2-x_3+x_4-3x_5=0\\ x_1+x_2-x_3\qquad\ +x_5=0\end{cases}$$

解空间（作为R^5的子空间）的一组标准正交基。

4-5 证试：任何一个二阶酉矩阵U可以分解为

$$U=\begin{pmatrix}e^{i\theta_1} & 0\\ 0 & e^{i\theta_2}\end{pmatrix}\begin{pmatrix}\cos\tau & \sin\tau\\ -\sin\tau & \cos\tau\end{pmatrix}\begin{pmatrix}e^{i\theta_3} & 0\\ 0 & e^{i\theta_4}\end{pmatrix}$$

其中θ_1，θ_2，θ_3，θ_4，τ都是实数。

4-6 将复矩阵U写为形如$U=P+iQ$，其中P、Q均为实矩阵，$i=\sqrt{-1}$，试证：U为酉矩阵的充要条件是$P^TP+Q^TQ=I$及P^TQ为对称矩阵。

4-7 设A、B是两个同阶正交矩阵，并且$|A|=-|B|$，试证：$A+B$是降秩的。

4-8 试求一个酉矩阵U，使矩阵

$$A=\begin{pmatrix}0 & 1 & 2\\ 1 & 0 & -1\\ 1 & -1 & 0\end{pmatrix}$$

酉相似于上三角矩阵。

4-9 设n阶酉矩阵U的特征根不等于-1，那么n阶矩阵$I+U$满秩，且

$$H=i(I-U)(I+U)^{-1}$$

是Hermite矩阵。反之，若H是Hermite矩阵，则n阶矩阵$I-iH$满秩，且

$$U=(I+iH)(I-iH)^{-1}$$

是酉矩阵。

4-10 若实矩阵S、T分别是实对称和实反对称矩阵，且$\det(I-T-iS)\neq0$，则$(I+T+iS)(I-T-iS)^{-1}$是酉矩阵。

4-11 试求酉矩阵U，使$U^{-1}AU$为对角矩阵。已知

（1） $$A=\begin{pmatrix}\frac{1}{3} & -\frac{1}{3\sqrt{2}} & -\frac{i}{\sqrt{6}}\\ -\frac{1}{3\sqrt{2}} & \frac{1}{6} & \frac{i}{2\sqrt{3}}\\ \frac{i}{\sqrt{6}} & -\frac{i}{2\sqrt{3}} & \frac{1}{2}\end{pmatrix}$$；（2） $$A=\begin{pmatrix}0 & i & 1\\ -i & 0 & 0\\ 1 & 0 & 0\end{pmatrix}$$

4-12 设 $A=(a_{ij})$ 和 $B=(b_{ij})$ 都是 n 阶正定实对称矩阵，试证

$$C=(a_{ij}b_{ij})$$

也是 n 阶正定矩阵。

4-13 设 A、B 是两个同阶正定 Hermite 矩阵，则 $|\lambda B-A|=0$ 的根全是实数。

4-14 设 A 是 n 阶实对称矩阵，证明：A 正定的充要条件是 A 的特征多项式的根全大于零。

4-15 设 A、B 都是实对称矩阵，证明：存在正交矩阵 T，使 $T^{-1}AT=B$ 的充要条件是 A、B 的特征多项式的根全部相同。

4-16 设 A 是 n 阶实对称矩阵，且 $A^2=A$，证明存在正交矩阵 T，使得

$$T^{-1}AT=\begin{pmatrix} I_r & \\ & 0 \end{pmatrix}$$

4-17 设 A 是 n 阶实对称矩阵，且 $A^2=I$，证明：存在正交矩阵 T，使得

$$T^{-1}AT=\begin{pmatrix} I_r & 0 \\ 0 & -I_{n-r} \end{pmatrix}$$

4-18 试将下列 Hermite 矩阵在复相合下化为标准形

（1）
$$\begin{pmatrix} 1 & 1+i & 2+i & \cdots & (n-1)+i \\ 1-i & 1 & 0 & \cdots & 0 \\ 2-i & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & \\ (n-1)-i & 0 & 0 & \cdots & 1 \end{pmatrix}$$

（2）
$$\begin{pmatrix} 1 & \frac{i}{2} & & & \\ -\frac{i}{2} & 1 & \frac{i}{2} & & \\ & -\frac{i}{2} & 1 & \ddots & \\ & & \ddots & 1 & \frac{i}{2} \\ & & & -\frac{i}{2} & 1 \end{pmatrix}$$

4-19 试将下列复对称矩阵化为标准形

（1）
$$\begin{pmatrix} 1 & 1+\mathrm{i} & \cdots & n+\mathrm{i} \\ 1+\mathrm{i} & 1 & 0 \cdots & 0 \\ 2+\mathrm{i} & 0 & 1 \cdots & 0 \\ \vdots & \vdots & \vdots \ddots & \vdots \\ n+\mathrm{i} & 0 & \cdots \cdots & 1 \end{pmatrix}$$

（2）
$$\begin{pmatrix} 1 & \frac{\mathrm{i}}{2} & & & \\ \frac{\mathrm{i}}{2} & 1 & \frac{\mathrm{i}}{2} & & \\ & \frac{\mathrm{i}}{2} & 1 & \ddots & \\ & & \ddots & 1 & \frac{\mathrm{i}}{2} \\ & & & \frac{\mathrm{i}}{2} & 1 \end{pmatrix}$$

4-20 设 $A=A^H\in C^{n\times n}$，则总存在 $t>0$，使得 $A+tI$ 是正定 Hermite 矩阵。

4-21 设 $A=A^H$，$B=B^H$，均正定，若 $AB=(AB)^H$，则 AB 正定。

4-22 设 $A=A^H$，$B=B^H$，均正定，举例说明 $\frac{1}{2}(AB+B^HA^H)$ 未必正定。

4-23 设 $A=(a_{ij})$，其中 $a_{ij}=\frac{1}{i+j-1}$，则 A 正定。

4-24 设 $A=A^H$，若有自然数 k，使得 $A^k=0$，则 $A=0$。

4-25 设 A 是正定 Hermite 矩阵，且 A 还是酉矩阵，则 $A=I$。

4-26 设 A 是复矩阵，试证：(1) A 可以唯一地写成 $A=S+\mathrm{i}T$，其中 S 和 T 都是 Hermite 矩阵。(2) A 可以唯一地写成 $A=B+C$，其中 B 是 Hermite 矩阵，C 是反 Hermite 矩阵。

第五章　正规矩阵与矩阵偶的标准形

前一章我们讨论了酉矩阵和 Hermite 矩阵，以及它们与对角矩阵相似的问题。这一章将研究一类更普遍的矩阵——正规矩阵以及它与对角矩阵相似的问题。此外还将研究两种矩阵偶的标准形。

§ 5.1　正规矩阵

定义 5.1.1　若 n 阶复矩阵 A 满足

$$AA^H = A^H A \tag{5-1}$$

便称 A 为**正规矩阵**。当 A 是实矩阵时，称满足

$$AA^T = A^T A$$

的矩阵 A 为**实正规矩阵**。

不难验证，对角矩阵、酉矩阵、Hermite 矩阵与反 Hermite 矩阵都是正规矩阵。正交矩阵、实对称矩阵与实反对称矩阵都是实正规矩阵。

引理 5.1.1　设 A 为正规矩阵，A 与 B 酉相似，则 B 是正规矩阵。

〔证明〕 存在酉矩阵 U，使得

$$B = U^{-1}AU$$

因此

$$\begin{aligned} BB^H &= (U^{-1}AU)(U^{-1}AU)^H \\ &= U^{-1}AA^HU = U^{-1}A^HAU \\ &= U^{-1}A^HUU^{-1}AU \\ &= (U^{-1}AU)^H(U^{-1}AU) = B^HB \end{aligned}$$

】

引理 5.1.2　设 A 是正规矩阵，且 A 是三角形矩阵，则 A 是对角矩阵。

〔证明〕 设 A 为上三角矩阵

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix}$$

根据正规矩阵定义得

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ & & a_{nn} \end{pmatrix} \begin{pmatrix} \bar{a}_{11} & & \\ \vdots & \ddots & \\ \bar{a}_{1n} & \cdots & \bar{a}_{nn} \end{pmatrix} = \begin{pmatrix} \bar{a}_{11} & & \\ \vdots & \ddots & \\ \bar{a}_{1n} & \cdots & \bar{a}_{nn} \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ & & a_{nn} \end{pmatrix} \tag{5-2}$$

将上式两端矩阵相乘，然后写出等号两端矩阵第一行第一列的元素得

$$a_{11}\bar{a}_{11}+a_{12}\bar{a}_{12}+\cdots+a_{1n}\bar{a}_{1n}=\bar{a}_{11}a_{11}$$

于是

$$a_{12}\bar{a}_{12}+\cdots+a_{1n}\bar{a}_{1n}=0$$

因为对于每一个 i，$a_{1i}\bar{a}_{1i}\geqslant 0$，故得

$$a_{1i}=0 \qquad (i=2,\ 3,\ \cdots,\ n)$$

再写出式(5-2)两端矩阵的第二行第二列得

$$a_{22}\bar{a}_{22}+a_{23}\bar{a}_{23}+\cdots+a_{2n}\bar{a}_{2n}=\bar{a}_{22}a_{22}$$

于是

$$a_{23}\bar{a}_{23}+\cdots+a_{2n}\bar{a}_{2n}=0$$

因此

$$a_{2i}=0 \qquad (i=3,\ 4,\ \cdots,\ n)$$

继续上述步骤便得

$$a_{ij}=0 \qquad (j>i)$$

此即 A 为对角矩阵。 ▌

定理 5.1.1 n 阶复矩阵 A 是正规矩阵的充要条件是 A 与对角矩阵酉相似，即存在 n 阶酉矩阵 U，使得

$$U^H AU=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n) \tag{5-3}$$

其中 λ_1，λ_2，…，λ_n 是 A 的 n 个特征值。

〔证明〕 *必要性* 设 A 是 n 阶复矩阵，根据定理 4.3.3，存在 n 阶酉矩阵 U，使得

$$B=U^H AU=U^{-1}AU$$

其中 B 为三角形矩阵，它的对角线上元素恰是 A 的特征值。

由引理 5.1.1 与引理 5.1.2 知，B 是对角矩阵，即

$$U^{-1}AU=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)=B$$

充分性 由引理 5.1.1 立即证得。 ▌

推论 5.1.1 设 A 是 n 阶正规矩阵，那么 A 与 A^H 可以同时化为对角矩阵。即存在 n 阶酉矩阵 U，使得 $U^H AU$ 与 $U^H A^H U$ 都是对角矩阵。

〔证明〕 因为

$$U^H AU=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)$$

所以

$$U^H A^H U=\mathrm{diag}(\bar{\lambda}_1,\ \bar{\lambda}_2,\ \cdots,\ \bar{\lambda}_n)$$ ∎

推论 5.1.2 设 n 阶正规矩阵 A 的特征值为 $\lambda_1,\ \cdots,\ \lambda_n$，则 A^H 的 n 个特征值为 $\bar{\lambda}_1,\ \cdots,\ \bar{\lambda}_n$。

我们知道，酉矩阵，Hermite 矩阵和反 Hermite 矩阵都是正规矩阵。下面的定理给出一个正规矩阵是酉矩阵，Hermite 矩阵和反 Hermite 矩阵的充要条件。

定理 5.1.2 设 A 是正规矩阵，则

(1) A 为 Hermite 矩阵的充要条件是 A 的特征值全为实数。

(2) A 为反 Hermite 矩阵的充要条件是 A 的所有特征值的实部均为零。

(3) A 为酉矩阵的充要条件是 A 的所有特征值的模全为 1。

〔证明〕 对于正规矩阵 A，存在酉矩阵 U，使得

$$U^H AU=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)=\Lambda$$

(1) 当 $A=A^H$ 时，$\Lambda=\Lambda^H$，因此 $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$ 全为实数。

相反地，当 $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$ 全为实数时，就有 $\Lambda=\Lambda^H$，所以 $A=A^H$。

(2) 当 $A=-A^H$ 时，$\Lambda=-\Lambda^H$，因此 $\lambda_i=-\bar{\lambda}_i$，这说明 λ_i 的实部为零。

相反地，若 $\lambda_j=\mathrm{i}\beta_j,\ \beta_j\in R\quad(j=1,2,\cdots,\ n)$则 $\bar{\lambda}_j=-\mathrm{i}\beta_j=-\lambda_j$，所以 $\bar{\Lambda}=-\Lambda$，因此 $A=-A^H$。

(3) 当 $AA^H=I$ 时，$\Lambda\Lambda^H=I$，因此 $\lambda_i\bar{\lambda}_i=1$，此即 $|\lambda_i|=1$。

相反地，若 $|\lambda_i|=1$，即 $\lambda_i\bar{\lambda}_i=1$，故 $\Lambda=\Lambda^H=I$，因此 $AA^H=I$。 ∎

推论 5.1.3 若 A 为 n 阶反 Hermite 矩阵，则 A 的特征值有 k 对共轭虚数 $\pm\mathrm{i}\beta_j$，其余的 $(n-2k)$ 个特征值全为零，且 $\mathrm{rank}A=2k$。

最后我们来讨论正规矩阵的谱分解。

引理 5.1.3 设 E 为 n 阶复矩阵，则 $E^2=E=E^H$ 的充要条件是存在 $n\times r$ 矩阵 U，使得

$$E=UU^H$$

其中 $r=\mathrm{rank}(E)$，U 的 r 个列向量是两两正交的单位向量，记 $U\in U_r^{n\times r}$。

〔证明〕 *必要性* 设 $r=\mathrm{rank}(E)$，故它有 r 个线性无关的列向量，从这 r 个列向量出发，用 Schmidt 方法作出 r 个两两正交的单位向量，以这 r 个向量为列构成一个 $n\times r$ 矩阵 $U\in U_r^{n\times r}$。E 的 n 个列向量都可以由 U 的 r 个列向量线性表出。即若 $U=(\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \cdots,\ \boldsymbol{e}_r)\in U_r^{n\times r}$，$E=(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_n)$则

$$\boldsymbol{x}_i=c_{i1}\boldsymbol{e}_1+c_{i2}\boldsymbol{e}_2+\cdots+c_{ir}\boldsymbol{e}_r\qquad(i=1,\ 2,\ \cdots,\ n)$$

所以

$$E=(x_1,\ x_2,\ \cdots,\ x_n)=(e_1,\ e_2,\ \cdots e_r)\begin{pmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ c_{12} & c_{22} & \cdots & c_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ c_{1r} & c_{2r} & \cdots & c_{nr} \end{pmatrix}$$

$$=UV^H$$

其中

$$V=\begin{pmatrix} \bar{c}_{11} & \bar{c}_{12} & \cdots & \bar{c}_{1r} \\ \bar{c}_{21} & \bar{c}_{22} & \cdots & \bar{c}_{2r} \\ \cdots & \cdots & \cdots & \cdots \\ \bar{c}_{n1} & \bar{c}_{n2} & \cdots & \bar{c}_{nr} \end{pmatrix}\in C^{n\times r}$$

由于向量组 $x_1,\ x_2,\ \cdots,\ x_n$ 的秩为 r，所以 V^H 的秩为 r。

下边证明 $V=U$。

事实上，由 $E=E^H=E^2$ 得

$$UV^H=VU^H=VU^HUV^H$$

注意到 $U^HU=I_r$，所以

$$UV^H=VV^H$$

或

$$(U-V)V^H=0$$

因为 V^H 的秩为 r，于是 $U-V$ 的秩为 0，因此

$$U-V=0$$

于是

$$E=UU^H$$

充分性　若 $E=UU^H$

则

$$E^2=E=E^H$$

设 A 是 n 阶正规矩阵，由定理 5.1.1 有

$$A=U\Lambda U^H=U\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)U^H$$

命

$$U=(\boldsymbol{X}_1,\ \boldsymbol{X}_2,\ \cdots,\ \boldsymbol{X}_n)$$

则

$$A=(X_1, X_2, \cdots, X_n)\operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)\begin{pmatrix} X_1^H \\ X_2^H \\ \vdots \\ X_n^H \end{pmatrix}$$

$$=\lambda_1 X_1 X_1^H+\lambda_2 X_2 X_2^H+\cdots+\lambda_n X_n X_n^H$$

命

$$G_i=X_i X_i^H \tag{5-4}$$

显然，G_i 是 n 阶矩阵，则

$$A=\lambda_1 G_1+\lambda_2 G_2+\cdots+\lambda_n G_n \tag{5-5}$$

式(5-5)称为正规矩阵的谱分解。

当正规矩阵 A 有重根时，还可以把式(5-5)右端进一步简化。设 A 有 r 个相异的特征值 λ_1，λ_2，…，λ_r，其重数分别为 n_1，n_2，…，n_r，则从式(5-5)得

$$A=\sum_{j=1}^{r}\lambda_j\left(\sum_{i=1}^{n_j}X_{ji}X_{ji}^H\right)$$

命

$$E_j=\sum_{i=1}^{n_j}X_{ji}X_{ji}^H \tag{5-6}$$

则

$$A=\sum_{j=1}^{r}\lambda_j E_j \tag{5-7}$$

不难验证

$$E_j=E_j^H=E_j^2$$

定理 5.1.3 设 n 阶复矩阵 A 有 r 个相异的特征值 λ_1，λ_2，…，λ_r，则 A 为正规矩阵的充要条件是存在 r 个矩阵 E_1，E_2，…，E_r，使得

(1) $A=\sum_{j=1}^{r}\lambda_j E_j$

(2) $E_j^2=E_j=E_j^H$

(3) $E_j E_k=0 \quad (j\neq k)$

(4) $\sum_{j=1}^{r}E_j=I$

(5) 满足上述性质的 E_j 是唯一的。

(6) $\operatorname{rank}(E_j)=n_j$

〔证明〕 必要性 (1)已证明式(5-5)。

(2)-(3)的证明作为练习留给读者。

(4)若命

$$U_j=(X_{j1}, X_{j2}, \cdots, X_{jn_j})$$

则由式(5-6)，知

$$E_j=U_jU_j^H$$

又因

$$U=(U_1, U_2, \cdots, U_r)$$

故有

$$I=UU^H=\sum_{j=1}^{r}U_jU_j^H=\sum_{j=1}^{r}E_j$$

(5) 按(1)、(2)、(3)，有

$$E_jA=E_j\left(\sum_{i=1}^{r}\lambda_iE_i\right)=\lambda_jE_j^2=\lambda_jE_j$$

$$=\left(\sum_{i=1}^{r}\lambda_iE_i\right)E_j=AE_j$$

此即

$$E_jA=\lambda_jE_j=AE_j \tag{5-8}$$

若还有 $\tilde{E}_i$ 满足上述条件，即

$$\tilde{E}_iA=\lambda_i\tilde{E}_i=A\tilde{E}_i \tag{5-9}$$

将 $\tilde{E}_i$ 右乘式(5-8)减去 E_j 左乘式(5-9)得

$$(\lambda_i-\lambda_j)E_j\tilde{E}_i=0$$

因为 $\lambda_i\neq\lambda_j$，故得

$$E_j\tilde{E}_i=0$$

从而有

$$E_j=E_jI=E_j\left(\sum_{i=1}^{r}\tilde{E}_i\right)=E_j\tilde{E}_j$$

$$=\left(\sum_{i=1}^{r}E_i\right)\tilde{E}_j=I\tilde{E}_j=\tilde{E}_j$$

(6) 由式(5-6)

$$E_j=(X_{j1}, X_{j2}, \cdots, X_{jn_j})\begin{pmatrix}X_{j1}^H\\ \vdots\\ X_{jn_j}^H\end{pmatrix}$$

因为矩阵$(X_{j1}, X_{j2}, \cdots, X_{jn_j})$的秩为$n_j$，所以

$$\text{rank}(E_j) \leqslant n_j$$

又按性质(4)知

$$I = \sum_{j=1}^{r} E_j$$

故

$$\begin{aligned}\text{rank}(I) = n &\leqslant \text{秩}\ E_1 + \cdots + \text{秩}\ E_r \\ &\leqslant n_1 + \cdots + n_r\end{aligned}$$

但是

$$n = n_1 + \cdots + n_r$$

故

$$\text{rank}(E_j) = n_j \quad (j = 1, 2, \cdots, r)$$

充分性　设A满足性质(1)—(6)，则

$$\begin{aligned}AA^H &= \left(\sum_{i=1}^{r} \lambda_i E_i\right)\left(\sum_{j=1}^{r} \lambda_j E_j\right) \\ &= \left(\sum_{i=1}^{r} \lambda_i E_i\right)\left(\sum_{j=1}^{r} \bar{\lambda}_j E_j^H\right) \\ &= \sum_{i=1}^{r} \lambda_i \bar{\lambda}_i E_i\end{aligned}$$

类似地可得

$$A^H A = \sum_{i=1}^{r} \lambda_i \bar{\lambda}_i E_i$$

所以A为正规矩阵。 ▌

§ 5.2　实正规矩阵在正交相似下的标准形

由定理5.1.1知，任何一个实正规矩阵必酉相似于一个对角矩阵，即实正规矩阵在酉相似下的标准形是对角矩阵。本节研究实正规矩阵在正交相似下的标准形。

下面先来证明任意实矩阵正交相似于准三角矩阵，它是Schur定理推广到实矩阵的结果。

引理5.2.1　设$\alpha_1 = (a_{11}, a_{21}, \cdots, a_{n1})^T$是模为1的$n$元实向量，那么存在正交矩阵$Q$以$\alpha_1$为它的第i列向量（$i = 1, 2, \cdots, n$）。

〔证明〕　仿照引理3.2.8的证明。考虑线性齐次方程

$$\boldsymbol{a}_1^T\boldsymbol{x}=a_{11}x_1+a_{21}x_2+\cdots+a_{n1}x_n=0$$

其中列向量 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^T$。方程的任何一个解向量 $\boldsymbol{x}$ 与 $\boldsymbol{a}_1$ 正交。设 $\boldsymbol{a}_2$ 为方程的某一个解向量，并且是单位向量，则 $\boldsymbol{a}_2^T\boldsymbol{a}_1=0$，$\boldsymbol{a}_2^T\boldsymbol{a}_2=1$

再解由两个方程构成的方程组

$$\boldsymbol{a}_1^T\boldsymbol{x}=0, \quad \boldsymbol{a}_2^T\boldsymbol{x}=0$$

取方程组的任何一个非零单位解向量 $\boldsymbol{a}_3$，显然 $\boldsymbol{a}_3$ 分别与 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$ 正交。

继续下去，假定已经求得了 $n-1$ 个两两正交的单位向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_{n-1}$，再解线性齐次方程组

$$\boldsymbol{a}_1^T\boldsymbol{x}=0, \quad \boldsymbol{a}_2^T\boldsymbol{x}=0, \quad \cdots, \quad \boldsymbol{a}_{n-1}^T\boldsymbol{x}=0$$

它有非零解，取非零单位解向量 $\boldsymbol{a}_n$。以 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ 为列组合而成的矩阵是正交矩阵。例如，$Q_1=(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n)$是以 $\boldsymbol{a}_1$ 为第一列的正交矩阵，$Q_n=(\boldsymbol{a}_2, \boldsymbol{a}_3, \cdots, \boldsymbol{a}_n, \boldsymbol{a}_1)$是以 $\boldsymbol{a}_1$ 为第 n 列的正交矩阵。其余类推。 ∎

定理 5.2.1 若 n 阶实矩阵 A 的实特征值为 $\lambda_{2s+1}, \lambda_{2s+2}, \cdots, \lambda_n$，复共轭特征值有 s 对，$a_1\pm ib_1, a_2\pm ib_2, \cdots, a_s\pm ib_s$，$(2s+k=n)$，则有 n 阶正交矩阵 Q，使得

$$Q^TAQ=\begin{pmatrix} T_1 & 0 \\ T_{21} & T_2 \end{pmatrix}=T \tag{5-10}$$

其中

$$T_1=\begin{pmatrix} A_1 & & & \\ A_{21} & A_2 & & \\ \cdots\cdots & \cdots\cdots & \ddots & \\ A_{s1} & A_{s2} & & A_s \end{pmatrix}, \quad T_2=\begin{pmatrix} \lambda_{2s+1} & & 0 \\ & \ddots & \\ * & & \lambda_n \end{pmatrix} \tag{5-11}$$

$A_1, A_2, \cdots, A_s$ 都是二阶矩阵，它们的特征值分别是 $a_1\pm ib_1, \cdots, a_s\pm ib_s$。

特别地，当 A 的特征值都是实数时，方阵 A 正交相似于一个下三角矩阵。

〔证明〕 用数学归纳法。

阶数为 1 时定理显然成立。现设 A 的阶数 $\leqslant m-1$ 定理成立，考虑阶数为 m 时的情况。下面分两种情况讨论。

(1) 设 A 有一个实特征值 λ_m，$\boldsymbol{a}_m$ 为与它对应的单位实特征列向量，由引理 5.2.1 知存在正交矩阵 Q_1 以 $\boldsymbol{a}_m$ 为它的第 m 个列向量，即 $Q_1=(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m)$，因此

$$Q_1^{-1}AQ_1=\begin{pmatrix} a'_{11} & a'_{12} & \cdots & a'_{1m-1} & 0 \\ & & & & \vdots \\ & & A_1 & & \lambda_m \end{pmatrix}$$

其中 A_1 是 $m-1$ 阶实矩阵，按归纳假设，存在 $m-1$ 阶正交矩阵 V_1，使得

$$
V_1^{-1}A_1V_1=\begin{pmatrix}
A_1 & & & & & & \\
A_{21} & A_2 & & & & & \\
\cdots\cdots & & \ddots & & & & \\
A_{s1} & \cdots\cdots & & A_s & & 0 & \\
 & & & & \lambda_{2s+1} & & \\
 & * & & & * & \ddots & \\
 & & & & & & \lambda_{m-1}
\end{pmatrix}
$$

命

$$
Q=Q_1\begin{pmatrix} V_1 & 0 \\ 0 & 1 \end{pmatrix}
$$

显然 Q 是 m 阶正交矩阵，且

$$
Q^{-1}AQ=\begin{pmatrix}
A_1 & & & & & \\
\cdots\cdots & \ddots & & & 0 & \\
A_{s1}\cdots\cdots & & A_s & & & \\
 & & & \lambda_{2s+1} & & \\
 & * & * & & \ddots & \\
 & & & & & \lambda_m
\end{pmatrix}=T
$$

T 的特征多项式（等于 A 的特征多项式）是

$$
\det(\lambda I_m-T)=\det(\lambda I_2-A_1)\det(\lambda I_2-A_2)\cdots
$$

$$
\det(\lambda I_2-A_s)(\lambda-\lambda_{2s+1})\cdots(\lambda-\lambda_m)
$$

此即矩阵 A（或 T）的实特征值为 λ_{2s+1}，…，λ_m，复特征值是 $a_1\pm ib_1$，$a_2\pm ib_2$，…，$a_s\pm ib_s$，其中 $a_j\pm ib_j$ 是二阶矩阵 A_j 的一对共轭复特征值，这就证明了定理。

（2） 设 A 的特征值都是复数，这时 $m=2s$。任取一个复特征值 $\lambda_s=a_s+ib_s$，其对应的特征向量（是复向量）$\boldsymbol{\alpha}_s$ 写为 $\boldsymbol{\alpha}_s=\boldsymbol{x}_s+i\boldsymbol{y}_s$。其中 $\boldsymbol{x}_s$ 与 $\boldsymbol{y}_s$ 都是实向量。根据 $A\boldsymbol{\alpha}_s=\lambda_s\boldsymbol{\alpha}_s$ 可得

$$
\begin{cases}
A\boldsymbol{x}_s=a_s\boldsymbol{x}_s-b_s\boldsymbol{y}_s \\
A\boldsymbol{y}_s=b_s\boldsymbol{x}_s+a_s\boldsymbol{y}_s
\end{cases}
\tag{5-12}
$$

现在证明实向量 $\boldsymbol{x}_s$ 与 $\boldsymbol{y}_s$ 线性无关。

首先，由于 $\boldsymbol{\alpha}_s\neq\boldsymbol{0}$ 知，$\boldsymbol{x}_s\neq\boldsymbol{0}$ 与 $\boldsymbol{y}_s\neq\boldsymbol{0}$ 同时成立。否则，若 $\boldsymbol{y}_s=0$，则根据（5-12）便

得 $A_s\boldsymbol{x}_s=a\boldsymbol{x}_s$，这表明 A 有实特征值 a_s，与假设矛盾，所以 $\boldsymbol{y}_s\neq 0$。同理 $\boldsymbol{x}_s\neq 0$。其次，若 $\boldsymbol{x}_s$ 与 $\boldsymbol{y}_s$ 线性相关，便有 $\boldsymbol{x}_s=k\boldsymbol{y}_s$，这时再考虑到式(5-12)便有

$$A\boldsymbol{y}_s=b_s\boldsymbol{x}_s+a_s\boldsymbol{y}_s=(b_sk+a_s)\boldsymbol{y}_s$$

这表明 A 有实特征值，又导出矛盾，这就证明了 $\boldsymbol{x}_s$ 与 $\boldsymbol{y}_s$ 线性无关。

由 $\boldsymbol{x}_s$ 与 $\boldsymbol{y}_s$ 应用 Schmidt 方法，得到

$$\begin{cases}\widetilde{\boldsymbol{x}}_s=b_{11}\boldsymbol{x}_s\\ \widetilde{\boldsymbol{y}}_s=b_{21}\boldsymbol{x}_s+b_{22}\boldsymbol{y}_s\end{cases} \tag{5-13}$$

其中 $b_{11}>0$，$b_{22}>0$，$\widetilde{\boldsymbol{y}}_s$ 与 $\widetilde{\boldsymbol{x}}_s$ 是互相正交的单位实向量。且由(5-12)与(5-13)得到

$$\begin{aligned}\begin{pmatrix}A\widetilde{\boldsymbol{x}}_s\\ A\widetilde{\boldsymbol{y}}_s\end{pmatrix}&=\begin{pmatrix}b_{11}&0\\ b_{21}&b_{22}\end{pmatrix}\begin{pmatrix}A\boldsymbol{x}_s\\ A\boldsymbol{y}_s\end{pmatrix}\\ &=\begin{pmatrix}b_{11}&0\\ b_{21}&b_{22}\end{pmatrix}\begin{pmatrix}a_s&-b_s\\ b_s&a_s\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_s\\ \boldsymbol{y}_s\end{pmatrix}\\ &=\begin{pmatrix}b_{11}&0\\ b_{21}&b_{22}\end{pmatrix}\begin{pmatrix}a_s&-b_s\\ b_s&a_s\end{pmatrix}\begin{pmatrix}b_{11}&0\\ b_{21}&b_{22}\end{pmatrix}^{-1}\begin{pmatrix}\widetilde{\boldsymbol{x}}_s\\ \widetilde{\boldsymbol{y}}_s\end{pmatrix}\\ &=\begin{pmatrix}\tilde{a}_{11}&\tilde{a}_{12}\\ \tilde{a}_{12}&\tilde{a}_{22}\end{pmatrix}\begin{pmatrix}\widetilde{\boldsymbol{x}}_s\\ \widetilde{\boldsymbol{y}}_s\end{pmatrix}\end{aligned} \tag{5-14}$$

其中

$$\begin{pmatrix}\tilde{a}_{11}&\tilde{a}_{12}\\ \tilde{a}_{21}&\tilde{a}_{22}\end{pmatrix}=\begin{pmatrix}b_{11}&0\\ b_{21}&b_{22}\end{pmatrix}\begin{pmatrix}a_s&-b_s\\ b_s&a_s\end{pmatrix}\begin{pmatrix}b_{11}&0\\ b_{21}&b_{22}\end{pmatrix}^{-1} \tag{5-15}$$

根据引理 5.2.1 不难得到，存在一个 m 阶正交矩阵

$$Q_1=(V,\ \widetilde{\boldsymbol{x}}_s,\ \widetilde{\boldsymbol{y}}_s)$$

其中 V 是 $m\times(m-2)$ 矩阵，且由 Q_1 的正交性得

$$V^T\widetilde{\boldsymbol{x}}_s=V^T\widetilde{\boldsymbol{y}}_s=0 \tag{5-16}$$

因此

$$Q_1^{-1}AQ_1=\begin{pmatrix}V^TAV & V^TA\widetilde{\boldsymbol{x}}_s & V^TA\widetilde{\boldsymbol{y}}_s\\ \widetilde{\boldsymbol{x}}_s^TAV & \widetilde{\boldsymbol{x}}_s^TA\widetilde{\boldsymbol{x}}_s & \widetilde{\boldsymbol{x}}_s^TA\widetilde{\boldsymbol{y}}_s\\ \widetilde{\boldsymbol{y}}_s^TAV & \widetilde{\boldsymbol{y}}_s^TA\widetilde{\boldsymbol{x}}_s & \widetilde{\boldsymbol{y}}_s^TA\widetilde{\boldsymbol{y}}_s\end{pmatrix} \tag{5-17}$$

根据式(5-14)与式(5-16)得到

$$\begin{cases}\widetilde{x}_s^T A\ \widetilde{x}_s=\tilde{a}_{11}, & \widetilde{y}_s^T A\widetilde{x}_s=\tilde{a}_{12}\\ \widetilde{x}_s^T A\ \widetilde{y}_s=\tilde{a}_{21}, & \widetilde{y}_s A\widetilde{y}_s=\tilde{a}_{22}\\ V^T A\ \widetilde{x}_s=V^T A\widetilde{y}_s=0\end{cases} \tag{5-18}$$

把式(5-18)代入式(5-17)得到

$$Q_1^{-1}AQ_1=\begin{pmatrix}V^TAV & 0 & 0\\ \widetilde{x}_s^TAV & \tilde{a}_{11} & \tilde{a}_{21}\\ \widetilde{y}_s^TAV & \tilde{a}_{12} & \tilde{a}_{22}\end{pmatrix} \tag{5-19}$$

由式(5-15)知

$$A_s=\begin{pmatrix}\tilde{a}_{11} & \tilde{a}_{21}\\ \tilde{a}_{12} & \tilde{a}_{22}\end{pmatrix}\quad 与\quad \begin{pmatrix}a_s & -b_s\\ b_s & a_s\end{pmatrix}$$

的特征值相同，而后者的特征值为

$$\lambda_s=a_s+\mathrm{i}b_s\quad 与\quad \bar{\lambda}_s=a_s-\mathrm{i}b_s$$

在式(5-19)中 $V^TA\ V$ 是 $m-2$ 阶实矩阵，根据归纳假设存在 $m-2$ 阶正交矩阵 Q_2，使得

$$Q_2^{-1}V^TA\ VQ_2=\begin{pmatrix}A_1 & & & 0\\ A_{21} & A_2 & & \\ \vdots & & \ddots & \\ A_{s-1,1} & \cdots & \cdots & A_{s-1}\end{pmatrix}$$

命

$$Q=Q_1\begin{pmatrix}Q_2 & \\ & I_2\end{pmatrix}$$

其中 I_2 为二阶单位矩阵，Q 是 m 阶正交矩阵。由直接计算可知

$$Q^{-1}A\ Q=\begin{pmatrix}A_1 & & & 0\\ A_{21} & A_2 & & \\ \vdots & & \ddots & \\ A_{s1} & \cdots & \cdots & A_s\end{pmatrix}$$

因为 $\det(\lambda I-A)=\det(\lambda I_2-A_1)\det(\lambda I_2-A_2)\cdots\det(\lambda I_2-A_s)$，所以 A 的 s 对共轭复特征值为 $a_1\pm \mathrm{i}b_1$，$a_2\pm \mathrm{i}b_2$，…，$a_s\pm \mathrm{i}b_s$，其中 $a_j\pm \mathrm{i}b_j$ 为二阶矩阵 A_j 的一对共轭复特征值。]

用类似的方法可以证明 A 正交相似于准上三角矩阵，即

推论 5.2.1 若 n 阶实矩阵 A 的实特征值为 λ_{2s+1}，λ_{2s+2}，…，λ_{2s+k}，复共轭特征值有 s 对，为 $a_1 \pm ib_1$，$a_2 \pm ib_2$，…，$a_s \pm ib_s (2s+k=n)$，则有 n 阶正交矩阵 Q，使得

$$Q^T A Q = \begin{pmatrix} T_1 & T_{12} \\ & T_2 \end{pmatrix} = T \tag{5-20}$$

其中

$$T_1 = \begin{pmatrix} A_1 & A_{12} & \cdots & A_{1s} \\ & A_2 & \cdots & A_{2s} \\ & & \ddots & \vdots \\ & & & A_s \end{pmatrix}, \quad T_2 = \begin{pmatrix} \lambda_{2s+1} & & * \\ & \ddots & \\ 0 & & \lambda_{2s+k} \end{pmatrix}$$

A_1，A_2，…，A_s 都是二阶矩阵，它们的特征值分别是 $a_1 \pm ib_1, a_2 \pm ib_2, \cdots, a_s \pm ib_s$。

为了研究实正规矩阵正交相似下的标准形，先介绍下述两引理。

引理 5.2.2 设 A、B 都是 n 阶实矩阵，且 A 正交相似于 B，那么 A 是实正规矩阵的充要条件是 B 为实正规矩阵。

〔证明〕 因为 A 正交相似于 B，所以存在正交矩阵 Q，满足

$$Q^{-1} A Q = B \tag{5-21}$$

由于 Q 是正交矩阵，因此

$$Q^{-1} A^T Q = B^T \tag{5-22}$$

两式相乘可得

$$Q^{-1} A A^T Q = B B^T$$

$$Q^{-1} A^T A Q = B^T B$$

因此 $AA^T = A^T A$ 与 $BB^T = B^T B$ 等价。 ∎

引理 5.2.3 设 G 是二阶实矩阵，特征值为 $a \pm ib$，则 G 是实正规矩阵的充要条件是存在二阶正交矩阵 Q，使得

$$Q^T G Q = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \tag{5-23}$$

〔证明〕 必要性 设 G 为实正规矩阵，根据定理 5.1.1 存在酉矩阵 U，使得

$$U^{-1} G U = \begin{pmatrix} a+ib & 0 \\ 0 & a-ib \end{pmatrix}$$

记 $U = (\boldsymbol{x}_1, \boldsymbol{x}_2)$，$\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 分别是 G 的对应于 $a+ib$ 与 $a-ib$ 的模为 1 的特征向量。因为 G 是实矩阵，所以 $\boldsymbol{x}_2 = \bar{\boldsymbol{x}}_1$。命

$$\boldsymbol{y}_1=\frac{\boldsymbol{x}_1+\boldsymbol{x}_2}{\sqrt{2}},\quad \boldsymbol{y}_2=\frac{\boldsymbol{x}_1-\boldsymbol{x}_2}{\sqrt{2}\ \mathrm{i}} \tag{5-24}$$

显然 $\boldsymbol{y}_1$ 与 $\boldsymbol{y}_2$ 都是实向量。可以根据 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 的复正交性条件来证明 $\boldsymbol{y}_1$ 与 $\boldsymbol{y}_2$ 是模为 1 的互相正交的实向量。事实上，

$$\boldsymbol{x}_1^T\boldsymbol{x}_1=\overline{\boldsymbol{x}_1^H\overline{\boldsymbol{x}}_1}=\overline{\boldsymbol{x}_1^H\boldsymbol{x}_2}=0$$

$$\boldsymbol{x}_1^T\boldsymbol{x}_2=\overline{\boldsymbol{x}_1^H\overline{\boldsymbol{x}}_2}=\overline{\boldsymbol{x}_1^H\boldsymbol{x}_1}=1$$

同样可得

$$\boldsymbol{x}_2^T\boldsymbol{x}_1=1,\quad \boldsymbol{x}_2^T\boldsymbol{x}_2=0$$

因此

$$\boldsymbol{y}_1^T\boldsymbol{y}_1=\frac{\boldsymbol{x}_1^T+\boldsymbol{x}_2^T}{\sqrt{2}}\ \frac{\boldsymbol{x}_1^T+\boldsymbol{x}_2}{\sqrt{2}}=1 \tag{5-25}$$

$$\boldsymbol{y}_1^T\boldsymbol{y}_2=\frac{\boldsymbol{x}_1^T+\boldsymbol{x}_2^T}{\sqrt{2}}\ \frac{\boldsymbol{x}_1-\boldsymbol{x}_2}{\sqrt{2}\ \mathrm{i}}=0 \tag{5-26}$$

$$\boldsymbol{y}_2^T\boldsymbol{y}_2=\frac{\boldsymbol{x}_1^T-\boldsymbol{x}_2^T}{\sqrt{2}\ \mathrm{i}}\ \frac{\boldsymbol{x}_1-\boldsymbol{x}_2}{\sqrt{2}\ \mathrm{i}}=1 \tag{5-27}$$

若令 $Q=[\boldsymbol{y}_1,\boldsymbol{y}_2]$，显然它是正交矩阵，且

$$Q^TGQ=\begin{pmatrix}\boldsymbol{y}_1^T\\ \boldsymbol{y}_2^T\end{pmatrix}G[\boldsymbol{y}_1,\boldsymbol{y}_2]=\begin{pmatrix}\boldsymbol{y}_1^TG\boldsymbol{y}_1 & \boldsymbol{y}_1^TG\boldsymbol{y}_2\\ \boldsymbol{y}_2^TG\boldsymbol{y}_1 & \boldsymbol{y}_2^TG\boldsymbol{y}_2\end{pmatrix} \tag{5-28}$$

因为

$$G\boldsymbol{x}_1=(a+\mathrm{i}b)\boldsymbol{x}_1,\quad G\boldsymbol{x}_2=(a+\mathrm{i}b)\boldsymbol{x}_2$$

所以

$$\begin{cases}G\boldsymbol{y}_1=a\boldsymbol{y}_1-b\boldsymbol{y}_2\\ G\boldsymbol{y}_2=a\boldsymbol{y}_1+b\boldsymbol{y}_2\end{cases} \tag{5-29}$$

把(5-29)代入(5-28)式右端，考虑到(5-25),(5-26)，(5-27)三式便得

$$Q^TGQ=\begin{pmatrix}a & b\\ -b & a\end{pmatrix}$$

充分性　若存在正交矩阵 Q，使得

$$Q^TGQ=\begin{pmatrix}a & b\\ -b & a\end{pmatrix}$$

那么

$$GG^T = Q\begin{pmatrix} a & b \\ -b & a \end{pmatrix} Q^T Q \begin{pmatrix} a & -b \\ b & a \end{pmatrix} Q^T$$

$$= Q\begin{pmatrix} a & b \\ -b & a \end{pmatrix}\begin{pmatrix} a & -b \\ b & a \end{pmatrix} Q^T$$

$$= Q\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} a & b \\ -b & a \end{pmatrix} Q^T$$

$$= Q\begin{pmatrix} a & -b \\ b & a \end{pmatrix} Q^T Q \begin{pmatrix} a & b \\ -b & a \end{pmatrix} Q^T = G^T G$$ ▌

现在我们可以证明下面的定理。

定理 5.2.2 设n阶实矩阵A的实特征值为$\lambda_{2s+1}, \lambda_{2s+2}, \cdots, \lambda_{2s+k}$,复共轭特征值有$s$对,为$a_1 \pm ib_1$, $a_2 \pm ib_2$, …, $b_s \pm ib_s$($2s+k=n$)。则A是实正规矩阵的充要条件是存在n阶正交矩阵Q,使得

$$Q^T A Q = \mathrm{diag}(H_1,\ H_2,\ \cdots,\ H_s,\ \lambda_{2s+1},\ \cdots \lambda_{2s+k}) \tag{5-30}$$

其中

$$H_j = \begin{pmatrix} a_j & b_j \\ -b_j & a_j \end{pmatrix} \qquad (j=1,\ 2,\ \cdots,\ s)$$

〔证明〕 必要性 设A是实正规矩阵,根据推论5.2.1,存在n阶正交矩阵Q_1使得

$$Q_1^T A Q_1 = G = \begin{pmatrix} G_{11} & G_{12} \\ 0 & G_{22} \end{pmatrix} \tag{5-31}$$

其中G_{22}是上三角矩阵,G_{11}是由二阶矩阵组成的块状上三角矩阵,即

$$G_{11} = \begin{pmatrix} A_{11} & A_{12} \cdots A_{1s} \\ & A_{22} \cdots A_{2s} \\ & \ddots \vdots \\ & A_{ss} \end{pmatrix}, \qquad G_{22} = \begin{pmatrix} \lambda_{2s+1} & & * \\ & \ddots & \\ & & \lambda_{2s+k} \end{pmatrix}$$

A_1, A_2, …, A_s是特征值分别为$a_1 \pm ib_1$, …, $a_s \pm ib_s$的二阶实矩阵。因为A是实正规矩阵,所以由引理5.2.2知G也是实正规矩阵。即

$$\begin{pmatrix} G_{11} & G_{12} \\ 0 & G_{22} \end{pmatrix}\begin{pmatrix} G_{11}^T & 0 \\ G_{12}^T & G_{22}^T \end{pmatrix} = \begin{pmatrix} G_{11}^T & 0 \\ G_{12}^T & G_{22}^T \end{pmatrix}\begin{pmatrix} G_{11} & G_{12} \\ 0 & G_{22} \end{pmatrix}$$

比较等式两边可得

$$
\begin{cases}
G_{11}G_{11}^T + G_{12}G_{12}^T = G_{11}^T G_{11} \\
G_{12}G_{22}^T = G_{11}^T G_{12} \\
G_{22}G_{12}^T = G_{12}^T G_{11} \\
G_{22}G_{22}^T = G_{12}^T G_{12} + G_{22}^T G_{22}
\end{cases}
\tag{5-32}
$$

在式(5-32)的第一个等式两边取矩阵的迹，因为 $\mathrm{tr}(G_{11}G_{11}^T)=\mathrm{tr}(G_{11}^T G_{11})$，所以 $\mathrm{tr}(G_{12}G_{12}^T)=0$。由直接计算可知 $\mathrm{tr}(G_{12}G_{12}^T)$是 G_{12} 的所有元素的平方和，因此 $G_{12}=0$。代入式(5-31)后得

$$G_{11}G_{11}^T = G_{11}^T G_{11}$$

$$G_{22}G_{22}^T = G_{22}^T G_{22}$$

$$G_{12} = 0$$

这表明 G_{11}，G_{22} 均是正规矩阵，由引理5.1.2知，G_{22} 是对角矩阵，所以

$$G_{22} = \mathrm{diag}(\lambda_{2s+1},\ \lambda_{2s+2},\ \cdots,\ \lambda_{2s+k})$$

由 $G_{11}^T G_{11} = G_{11}G_{11}^T$ 可得到 s 个矩阵等式

$$
\begin{cases}
A_{11}^T A_{11} = A_{11}A_{11}^T + A_{12}A_{12}^T + \cdots + A_{1s}A_{1s}^T \\
A_{12}^T A_{12} + A_{22}A_{22}^T = A_{22}A_{22}^T + A_{23}A_{23}^T + \cdots + A_{2s}A_{2s}^T \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
A_{1s}^T A_{1s} + A_{2s}^T A_{2s} + \cdots + A_{ss}^T A_{ss} = A_{ss}A_{ss}^T
\end{cases}
\tag{5-33}
$$

在式(5-33)的第一个等式两边取矩阵的迹得

$$\mathrm{tr}(A_{11}^T A_{11}) = \mathrm{tr}(A_{11}A_{11}^T) + \cdots + \mathrm{tr}(A_{1s}A_{1s}^T)$$

故

$$\mathrm{tr}(A_{12}A_{12}^T) + \cdots + \mathrm{tr}(A_{1s}A_{1s}^T) = 0$$

因为 $\mathrm{tr}(A_{1j}A_{1j}^T)$是矩阵 A_{1j} 的所有元素之平方和，从而

$$A_{1j} = 0 \qquad (j=2,\ 3,\ \cdots,\ s)$$

把它代入式(5-33)的第二个等式以后，类似于对式(5-33)第一个等式的讨论便得

$$A_{2j} = 0 \qquad (j=3,\ 4,\ \cdots,\ s)$$

继续这个过程，最后得到

$$A_{ij} = 0, \qquad (i=1,\ 2,\ \cdots,\ s, \qquad i<j\leqslant s)$$

因此

$$G_{11}=\begin{pmatrix} A_{11} & & & \\ & A_{22} & & \\ & & \ddots & \\ & & & A_{ss} \end{pmatrix}=\text{diag}(A_{11},\ A_{22},\ \cdots,\ A_{ss})$$

式中每一个 $A_{jj}(j=1,\ 2,\ \cdots,\ s)$都是实正规矩阵，其特征值为 $a_j \pm ib_j$。

把 G_{11} 与 G_{22} 的表达式代入式(5-31)得到

$$Q_1^T AQ_1=\text{diag}(A_{11},\ A_{22},\ \cdots,\ A_{ss},\ \lambda_{2s+1},\ \cdots,\ \lambda_{2s+k})$$

根据引理 5.2.3，对于每一个 A_{jj} 存在二阶正交矩阵 $\bar{Q}_j$，使得

$$\bar{Q}_j^T A_{jj}\bar{Q}_j=\begin{pmatrix} a_j & b_j \\ -b_j & a_j \end{pmatrix}=H_j$$

命

$$Q=Q_1\text{diag}(\bar{Q}_1,\ \bar{Q}_2,\ \cdots,\ \bar{Q}_s,\ I_k)$$

则

$$Q^T AQ=\text{diag}(H_1,\ H_2,\ \cdots,\ H_s,\ \lambda_{2s+1},\ \cdots,\ \lambda_{2s+k})$$

充分性　由式(5-30)得

$$A=Q\ \text{diag}(H_1,\ H_2,\ \cdots,\ H_s,\ \lambda_{2s+1},\ \cdots,\ \lambda_{2s+k})Q^T$$

因此直接计算得到

$$A^T A=AA^T$$

而 A 的特征值分布可由 A 与 $\text{diag}(H_1,\ \cdots,\ H_s,\ \lambda_{2s+1},\ \cdots,\ \lambda_{2s+k})$ 正交相似立刻得到。]

由于正交矩阵，实对称矩阵，实反对称矩阵都属于实正规矩阵，因此可得如下定理。

定理 5.2.3　n 阶实矩阵 A 为正交矩阵的充要条件是存在 n 阶正交矩阵 Q，使得

$$Q^T AQ=\text{diag}(H_1,\ H_2,\ \cdots,\ H_s,\ 1,\ \cdots,\ 1,\ -1,\ \cdots,\ -1)$$

其中

$$H_j=\begin{pmatrix} \cos\theta_j & \sin\theta_j \\ -\sin\theta_j & \cos\theta_j \end{pmatrix}\qquad (j=1,\ 2,\ \cdots,\ s)$$

〔证明〕 根据定理 5.2.2，我们只要判断正交矩阵的特征值的模为 1 即可，而这是上一章推论 4.2.1 的结论。现在利用定理 5.2.2 讨论正交矩阵 A 的特征值。

根据定理 5.2.2 知

$$Q^T AQ=\text{diag}(H_1,\ H_2,\ \cdots,\ H_s,\ \lambda_{2s+1},\ \cdots,\ \lambda_{2s+k})$$

因为 A 是正交矩阵等价于 $\text{diag}(H_1,\ H_2,\ \cdots,\ H_s,\ \lambda_{2s+1},\ \cdots,\ \lambda_{2s+k})$是正交矩阵，于是得到

$$H_j^T H_j = H_j H_j^T = I_2 \qquad (j=1, 2, \cdots, s)$$

$$\lambda_{2s+1}^2 = \lambda_{2s+2}^2 = \cdots = \lambda_{2s+k}^2$$

因此 A 的实特征值为 1 或 −1。而复特征值 $a_j \pm \mathrm{i} b_j$ 满足

$$\begin{pmatrix} a_j & b_j \\ -b_j & a_j \end{pmatrix} \begin{pmatrix} a_j & -b_j \\ b_j & a_j \end{pmatrix} = \begin{pmatrix} a_j & -b_j \\ b_j & a_j \end{pmatrix} \begin{pmatrix} a_j & b_j \\ -b_j & a_j \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

所以

$$a_j^2 + b_j^2 = 1 \qquad (j=1, 2, \cdots, s)$$

因此存在角度 θ_j 使得

$$\mathrm{e}^{\pm \mathrm{i}\theta_j} = \cos\theta_j \pm \mathrm{i}\sin\theta_j = a_j \pm \mathrm{i} b_j$$

即

$$H_j = \begin{pmatrix} \cos\theta_j & \sin\theta_j \\ -\sin\theta_j & \cos\theta_j \end{pmatrix} \qquad (j=1, 2, \cdots, s)$$

】

定理 5.2.4 n 阶实矩阵 A 为实对称矩阵的充要条件是存在 n 阶正交矩阵 Q，使得

$$Q^T A Q = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 全是实数。

〔证明〕 *方法一* 实对称矩阵的特征值全是实数，因此由定理 5.2.2 立刻证得。

方法二 根据定理 5.2.2 知

$$Q^T A Q = \mathrm{diag}(H_1, \cdots, H_s, \lambda_{2s+1}, \cdots, \lambda_{2s+k})$$

显然 A 的对称性等价于 $\mathrm{diag}(H_1, \cdots, H_s, \lambda_{2s+1}, \cdots, \lambda_{2s+k})$ 的对称性，故

$$H_j^T = H_j$$

所以

$$b_j = 0 \qquad (j=1, 2, \cdots, s)$$

这表明 A 的复特征值的虚部为零，所以实对称矩阵 A 的特征值全是实数。 】

定理 5.2.5 n 阶实矩阵 A 是反对称矩阵的充要条件是存在正交矩阵 Q，使得

$$Q^T A Q = \mathrm{diag}(H_1, H_2, \cdots, H_s, 0, \cdots, 0)$$

其中

$$H_j = \begin{pmatrix} 0 & b_j \\ -b_j & 0 \end{pmatrix} \qquad (j=1, 2, \cdots, s)$$

〔证明〕 由定理 5.2.2 得

$$Q^T A Q = \mathrm{diag}(H_1, H_2, \cdots, H_s, \lambda_{2s+1}, \cdots, \lambda_{2s+k})$$

因为 A 为反对称矩阵等价于 $\mathrm{diag}(H_1, H_2, \cdots, H_s, \lambda_{2s+1}, \cdots, \lambda_{2s+k})$ 为反对称矩阵，所以

$$H_j^T=-H_j, \quad \lambda_{2s+1}=\lambda_{2s+2}=\cdots=\lambda_{2s+k}=0$$

由 $H_j^T=-H_j$ 得

$$a_j=0 \qquad (j=1, 2, \cdots, s)$$

于是

$$H_j=\begin{pmatrix} 0 & b_j \\ -b_j & 0 \end{pmatrix} \qquad (j=1, 2, \cdots, s)$$

】

§5.3 反对称矩阵在相合下的标准形

对称（实和复）矩阵在相合下的标准形在§4.4介绍过，这节研究反对称矩阵在相合下的标准形。由于复反对称矩阵在相合下的标准形的讨论方法与实反对称矩阵在实相合下的标准形的完全一样，因此，下面只对实反对称矩阵的情况进行讨论。

定理 5.3.1 设 A 是 n 阶实反对称矩阵，则有 n 阶矩阵 P 使得

$$P^TAP=\mathrm{diag}\left(\begin{pmatrix} 0 & a_1 \\ -a_1 & 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 & a_s \\ -a_s & 0 \end{pmatrix}, 0\cdots, 0\right)$$

其中 P 的元素是矩阵 A 的元素的有理函数（即为两个多元多项式之商），且 $\det P=1$，而 $a_1, a_2, \cdots, a_s$ 都是非零实数，它们也是 A 的元素的有理函数。

〔证明〕 用归纳法。当 $n=1$ 时定理显然成立。

设 $n\leqslant m-1$ 时定理成立。现在来证 $n=m$ 时定理成立。当 $A=0$ 时，取 $P=I_n$ 即成。今设 $A\neq 0$，于是，存在元素 $a_{jk}\neq 0(j<k)$，作实相合

$$(P(2, k)P(1, j))^TA(P(2, k)P(1, k))$$

它是一个反对称矩阵，且第一行第二列的元素为 $a_{jk}\neq 0$。而 $P(2, k)P(1, j)$ 是行列式为1的常数矩阵，所以适合定理要求。因此不妨设 $a_{12}\neq 0$，把 A 写成分块矩阵

$$A=\begin{pmatrix} \begin{pmatrix} 0 & a_{12} \\ -a_{12} & 0 \end{pmatrix} & B_1 \\ -B_1^T & A_1 \end{pmatrix}$$

命

$$P_1=\begin{pmatrix} I_2 & \begin{pmatrix} 0 & a_{12}^{-1} \\ -a_{12}^{-1} & 0 \end{pmatrix}B_1 \\ 0 & I_{m-2} \end{pmatrix}$$

则

$$P_1^T A P_1=\begin{pmatrix} I_2 & 0 \\ B_1^T\begin{pmatrix}0 & -a_{12}^{-1}\\ a_{12}^{-1} & 0\end{pmatrix} & I_{m-2}\end{pmatrix}\begin{pmatrix}\begin{pmatrix}0 & a_{12}\\ -a_{12} & 0\end{pmatrix} & B_1 \\ -B_1^T & A_1\end{pmatrix}$$

$$\times\begin{pmatrix} I_2 & \begin{pmatrix}0 & a_{12}^{-1}\\ -a_{12}^{-1} & 0\end{pmatrix} B_1 \\ 0 & I_{m-2}\end{pmatrix}$$

$$=\begin{pmatrix}\begin{pmatrix}0 & a_{12}\\ -a_{12} & 0\end{pmatrix} & 0 \\ 0 & B_1^T\begin{pmatrix}0 & -a_{12}^{-1}\\ a_{12}^{-1} & 0\end{pmatrix}B_1+A_1\end{pmatrix}$$

不难验证 $\det P_1=1$，且 P_1 的元素是 A 的元素的有理函数。又 $m-2$ 阶矩阵

$$\bar{A}=B_1^T\begin{pmatrix}0 & -a_{12}^{-1}\\ a_{12}^{-1} & 0\end{pmatrix}B_1+A_1$$

是反对称矩阵。由归纳假设，存在 $m-2$ 阶矩阵 P_2，使得

$$P_2^T\bar{A}P_2=\mathrm{diag}\left\{\begin{pmatrix}0 & a_2\\ -a_2 & 0\end{pmatrix}, \cdots, \begin{pmatrix}0 & a_s\\ -a_s & 0\end{pmatrix}, 0, \cdots, 0\right\}$$

其中 P_2 适合 $\det P_2=1$，且 P_2 的元素是 $\bar{A}$ 的元素的有理函数，由于有理函数的有理函数仍为有理函数，所以 P_2 的元素是 A 的元素的有理函数。而 a_2，a_3，…，a_s 都是 A 的元素的有理函数。

命

$$P=P_1\begin{pmatrix}I_2 & 0\\ 0 & P_2\end{pmatrix}$$

显然 P 适合定理要求，$a_{12}=a_1$。这就证明了定理。 ∎

定理 5.3.2 设 A 为 n 阶反对称矩阵，那么存在 n 阶满秩矩阵 Q，使得

$$Q^T AQ=\mathrm{diag}\left\{\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}, \cdots, \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}, \underbrace{0, \cdots, 0}_{n-2s}\right\} \tag{5-34}$$

其中 A 的秩 $=2s$。

〔证明〕由定理 5.3.1 可知，存在满秩矩阵 P，使得

$$P^T AP=\mathrm{diag}\left\{\begin{pmatrix}0 & a_1\\ -a_1 & 0\end{pmatrix}, \cdots, \begin{pmatrix}0 & a_s\\ -a_s & 0\end{pmatrix}, 0, \cdots, 0\right\}$$

其中 $a_1a_2\cdots a_s\neq0$，显然 $\text{rank}(A)=2s$。

又因

$$\begin{pmatrix} 1 & 0 \\ -1 & a_j^{-1} \end{pmatrix}^T \begin{pmatrix} 0 & a_j \\ -a_j & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & a_j^{-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

所以命

$$Q=P\ \text{diag}\left(\begin{pmatrix} 1 & 0 \\ -1 & a_1^{-1} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ -1 & a_2^{-1} \end{pmatrix}, \cdots, \begin{pmatrix} 1 & 0 \\ -1 & a_s^{-1} \end{pmatrix}, 1, \cdots, 1\right)$$

则 Q^TAQ 便是所需之表达式。 ▮

式(5-34)称为反对称矩阵的相合标准形。

推论 5.3.1 反对称矩阵的秩为偶数。

§5.4 Hermite 矩阵偶在相合下的标准形

两个 Hermite 矩阵同时与对角矩阵复相合的问题是本节要研究的内容。在 Hermite 齐式问题上，就是两个 Hermite 齐式同时化简成标准形的问题。它在振动理论，物理及其它工程中有重要的应用。

定理 5.4.1 设 A、B 均为 n 阶 Hermite 矩阵，且 B 为正定的，则存在满秩的复矩阵 T，使得

$$T^HAT=\text{diag}(\mu_1, \mu_2, \cdots, \mu_n)=M \tag{5-35}$$

与

$$T^HBT=I \tag{5-36}$$

同时成立，其中 μ_1，μ_2，…，μ_n 是与 T 无关的实数。

〔证明〕 因为 B 是正定矩阵，所以存在复矩阵 T_1，使得

$$T_1^HBT_1=I$$

由于 $T_1^HAT_1$ 仍是 Hermite 矩阵，因此存在酉矩阵 T_2，使得

$$T_2^HT_1^HAT_1T_2=\text{diag}(\mu_1, \mu_2, \cdots, \mu_n)=M \tag{5-37}$$

其中 μ_1，μ_2，…，μ_n 是 Hermite 矩阵 $T_1^HAT_1$ 的实特征根。

若命 $T=T_1T_2$，便得

$$T^HAT=M$$

与

$$T^HBT=I$$

余下要证明 μ_1，μ_2，…，μ_n 与 T 无关。事实上，命

$$S=T_1^H A T_1 \quad (5-38)$$

则 μ_1, μ_2, …, μ_n 是特征方程

$$|\lambda I-S|=0 \quad (5-39)$$

的根。将式(5-37)与式(5-38)代入式(5-39)，有

$$|\lambda T_1^H B T_1 - T_1^H A T_1|=0$$

简化得

$$|\lambda B-A|=0 \quad (5-40)$$

反之，由式(5-40)可得式(5-39)。因此，μ_1, μ_2, …, μ_n 是 n 次代数方程(5-40)的根，由矩阵 A 与 B 所确定而与 T 无关。 ▌

若把定理 5.4.1 的结果应用到 Hermite 齐式上，便有如下定理。

定理5.4.2 设给定两个 Hermite 二次齐式

$$f_1=X^H A X=\sum_{i,j=1}^{n} a_{ij}\bar{x}_i x_j$$

$$f_2=X^H B X=\sum_{i,j=1}^{n} b_{ij}\bar{x}_i x_j$$

且 f_2 是正定的，则存在满秩的线性变换

$$X=TY$$

使得

$$f_1=\mu_1\bar{y}_1 y_1+\mu_2\bar{y}_2 y_2+\cdots+\mu_n\bar{y}_n y_n$$

$$f_2=\bar{y}_1 y_1+\bar{y}_2 y_2+\cdots+\bar{y}_n y_n$$

其中 μ_1, μ_2, …, μ_n 是方程 $|\lambda A-B|=0$ 的根（全是实的）。

在上述定理中出现的 μ_1, μ_2, …, μ_n 与矩阵 T，可以诱导出在计算和其他应用中有重要作用的两个矩阵的相对特征值和相对特征向量问题。现作简单介绍。

设 A、B 都是 n 阶 Hermite 矩阵，且 B 是正定的，求 λ 使方程

$$A\boldsymbol{x}=\lambda B\boldsymbol{x} \quad (5-41)$$

有非零解 $\boldsymbol{x}=(a_1, a_2, \cdots, a_n)^{\mathrm{T}}$。

式(5-41)有非零解的充要条件是关于 λ 的 n 次代数方程

$$|\lambda B-A|=0 \quad (5-42)$$

成立。称方程(5-42)是 **A 相对于 B 的特征方程**。它的根 λ_1, λ_2, …, λ_n 称为 **A 相对于 B 的广义特征值**，把 λ_i 代入式(5-41)所得非零解 $\boldsymbol{x}$ 称为**与 λ_i 相对应的广义特征向量**。

不难证明下述命题是正确的。

命题5.4.1 $A\boldsymbol{x}=\lambda B\boldsymbol{x}$ 的广义特征值与广义特征向量等价于 Hermite 矩阵 $S=T_1^H A T_1$ 的普通特征值与特征向量。

应用 Hermite 矩阵的性质可以证明下述定理。

定理 5.4.3 形如式(5-41)的广义特征值与广义特征向量有如下性质：

(1) 有 n 个实的广义特征值。

(2) 有 n 个线性无关的广义特征向量 $\boldsymbol{x}_1$, …, $\boldsymbol{x}_n$，即

$$A\boldsymbol{x}_k=\lambda_k B\boldsymbol{x}_k \qquad (k=1,\ 2,\ \cdots,\ n) \tag{5-43}$$

(3) 这 n 个广义特征向量可以这样选取，使其满足

$$(\boldsymbol{x}_i)^H B\boldsymbol{x}_j=\delta_{ij} \tag{5-44}$$

$$(\boldsymbol{x}_i)^H A\boldsymbol{x}_j=\lambda_j\delta_{ij} \tag{5-45}$$

注：$(\boldsymbol{x}_i)$表示把列向量 $\boldsymbol{x}$ 看成是列矩阵，故$(\boldsymbol{x}_i)^H$表示行向量，且它的分量是 $\boldsymbol{x}$ 的共轭分量。

〔证明〕 (1) 由定理 5.4.1 的证明过程可知。

(2) 设 $S=T_1^HAT_1$ 的 n 个特征值为 λ_1, λ_2, …, λ_n，它对应的线性无关的特征向量有 n 个，分别记为 $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, …, $\boldsymbol{y}_n$，故

$$S\boldsymbol{y}_k=\lambda_k\boldsymbol{y}_k \qquad (k=1,\ 2,\ \cdots,\ n)$$

或

$$T_1^HAT_1\boldsymbol{y}_k=\lambda_k\boldsymbol{y}_k \tag{5-46}$$

命

$$\boldsymbol{x}_k=T_1\boldsymbol{y}_k \tag{5-47}$$

代入式(5-46)得

$$T_1^HA\boldsymbol{x}_k=\lambda_kT_1^{-1}\boldsymbol{x}_k$$

或

$$A\boldsymbol{x}_k=\lambda_k(T_1^H)^{-1}T_1^{-1}\boldsymbol{x}_k$$

应用式(5-38)便得

$$A\boldsymbol{x}_k=\lambda_kB\boldsymbol{x}_k \tag{5-48}$$

这表明 $\boldsymbol{x}_k$ $(k=1,\ 2,\ \cdots,\ n)$是 n 个广义特征向量。现证明他们是线性无关的。

设

$$k_1\boldsymbol{x}_1+\cdots+k_n\boldsymbol{x}_k=\boldsymbol{0}$$

以式(5-47)代入上式得

$$k_1T_1\boldsymbol{y}_1+\cdots+k_nT_1\boldsymbol{y}_n=\boldsymbol{0}$$

以 T_1^{-1} 左乘上式每一项得

$$k\boldsymbol{y}_1+\cdots+k_n\boldsymbol{y}_n=\boldsymbol{0}$$

由于 y_1, y_2, …, y_n 是线性无关的，所以

$$k_1=k_2=\cdots=k_n=0$$

此即 x_1, x_2, …, x_n 是线性无关的。

(3) 根据 Hermite 矩阵与对角矩阵酉相似，便知可以这样选取 y_1, y_2, …, y_n 使它们满足

$$y_i^H y_j=\delta_{ij} \qquad (i, j=1, 2, \cdots, n)$$

由式(5-47)得

$$(T_1^{-1}x_i)^H(T_1^{-1}x_j)=\delta_{ij}$$

展开得

$$x_i^H(T_1^{-1})^H T_1^{-1}x_j=\delta_{ij}$$

由式(5-38)得

$$x_i^H B x_j=\delta_{ij}$$

又由式(5-48)，知

$$x_i^H A x_j=\lambda_j x_i^H B x_j=\lambda_j\delta_{ij}$$ ∎

在矩阵理论及其应用中，称满足式(5-44)的广义特征向量 x_1, x_2, …, x_n 为特征主向量。以这 n 个主向量为列向量构成的矩阵

$$T=(x_1, x_2, \cdots, x_n)$$

是满秩的，称为 A 相对于 B 的主矩阵。

定理 5.4.4 设 A、B 为 Hermite 矩阵，且 B 为正定的，则存在行列式等于 1 的矩阵 P，使

$$P^H AP=\mathrm{diag}(a_1, a_2, \cdots, a_n)$$

与

$$P^H BP=\mathrm{diag}(b_1, b_2, \cdots, b_n)$$

同时成立。其中 a_1, a_2, …, a_n 均为实数，b_1, b_2, …, b_n 均为正实数。

〔证明〕 由定理 5.4.1 知，存在满秩矩阵 T，使得

$$T^H AT=\mathrm{diag}(\mu_1, \mu_2, \cdots, \mu_n)$$

与

$$T^H BT=I$$

同时成立。命

$$P=\left(\frac{1}{\det T}\right)^{\frac{1}{n}}T$$

则

$$\det P=1$$

且

$$P^H A P=\left(\frac{1}{\det \bar{T}}\right)^{\frac{1}{n}} T^H A\left(\frac{1}{\det T}\right)^{\frac{1}{n}} T$$

$$=\left(\frac{1}{\det \bar{T}\det T}\right)^{\frac{1}{n}} T^H A T$$

$$=\mathrm{diag}(a_1,\ a_2,\ \cdots,\ a_n)$$

其中

$$a_i=\left(\frac{1}{\det \bar{T}\det T}\right)^{\frac{1}{n}}\mu_i \qquad (i=1,\ 2,\ \cdots,\ n)$$

$$P^H B P=\left(\frac{1}{\det \bar{T}\det T}\right)^{\frac{1}{n}} I=\mathrm{diag}(b_1,\ b_2,\ \cdots,\ b_n)$$

其中

$$b_1=b_2=\cdots=b_n=\left(\frac{1}{\det \bar{T}\det T}\right)^{\frac{1}{n}}>0$$

引理5.4.1 设a_i、$b_i(i=1,\ 2,\ \cdots,\ n)$均为正数，则有

$$(a_1a_2\cdots a_n)^{1/n}+(b_1b_2\cdots b_n)^{1/n}\leqslant[(a_1+b_1)(a_2+b_2)\cdots(a_n+b_n)]^{1/n}$$

当$b_i=ka_i$时等号成立。

〔证明〕 当$n=1$时，结论显然成立；当$n=2$时，有

$$[(a_1a_2)^{1/2}+(b_1b_2)^{1/2}]^2=a_1a_2+b_1b_2+2(a_1b_2a_2b_1)^{1/2}$$

$$\leqslant a_1a_2+b_1b_2+a_1b_2+a_2b_1=(a_1+b_1)(a_2+b_2)$$

即

$$(a_1a_2)^{1/2}+(b_1b_2)^{1/2}\leqslant[(a_1+b_1)(a_2+b_2)]^{1/2}$$

且当$a_1b_2=a_2b_1$时，等号成立。

设$n=k$时，结论成立，考虑$n=k+1$的情况：

$$\{(a_1+b_1)(a_2+b_2)\cdots(a_k+b_k)(a_{k+1}+b_{k+1})\}^{\frac{1}{k+1}}$$

$$=\{(a_1+b_1)(a_2+b_2)\cdots(a_k+b_k)\}^{\frac{1}{k}\cdot\frac{k}{k+1}}(a_{k+1}+b_{k+1})^{\frac{1}{k+1}}$$

$$\geqslant[(a_1a_2\cdots a_k)^{\frac{1}{k}}+(b_1b_2\cdots b_k)^{\frac{1}{k}}]^{\frac{k}{k+1}}(a_{k+1}+b_{k+1})^{\frac{1}{k+1}}$$

$$=\{[(a_1a_2\cdots a_k)^{\frac{1}{k}}+(b_1b_2\cdots b_k)^{\frac{1}{k}}]^k\}^{\frac{1}{k+1}}(a_{k+1}+b_{k+1})^{\frac{1}{k+1}}$$

$$\geqslant\{a_1a_2\cdots a_k+b_1b_2\cdots b_k\}^{\frac{1}{k+1}}(a_{k+1}+b_{k+1})^{\frac{1}{k+1}}$$

$$=\{(a_1a_2\cdots a_k+b_1b_2\cdots b_k)(a_{k+1}+b_{k+1})\}^{\frac{1}{k+1}}$$

$$=\{[(a_1a_2\cdots a_k+b_1b_2\cdots b_k)(a_{k+1}+b_{k+1})]^{\frac{1}{2}}\}^{\frac{2}{k+1}}$$

$$\geqslant\{(a_1a_2\cdots a_ka_{k+1})^{\frac{1}{2}}+(b_1b_2\cdots b_kb_{k+1})^{\frac{1}{2}}\}^{\frac{2}{k+1}}$$

$$=\{[(a_1a_2\cdots a_ka_{k+1})^{\frac{1}{2}}+(b_1b_2\cdots b_kb_{k+1})^{\frac{1}{2}}]^2\}^{\frac{1}{k+1}}$$

$$\geqslant\{a_1a_2\cdots a_ka_{k+1}+b_1b_2\cdots b_kb_{k+1}\}^{\frac{1}{k+1}}$$

此即证明了不等式在 $n=k+1$ 时也成立，且等号仅当 $b_i=ka_i$ 时成立。 】

定理5.4.5 设 A、B 为 n 阶正定 Hermite 矩阵，则

$$|A|^{\frac{1}{n}}+|B|^{\frac{1}{n}}\leqslant|A+B|^{\frac{1}{n}}$$

且等号在 $B=kA$ 时成立，k 为某实数。

〔证明〕 由定理 5.4.4 知有满秩矩阵 P，使得

$$|P|=1$$

$$P^HAP=\begin{pmatrix}a_1 & & \\ & \ddots & \\ & & a_n\end{pmatrix}$$

$$P^HBP=\begin{pmatrix}b_1 & & \\ & \ddots & \\ & & b_n\end{pmatrix}$$

由于 A、B 都是正定的，所以所有的 a_i 与 b_i 都是正数，并且

$$|A|=a_1a_2\cdots a_n,\quad |B|=b_1b_2\cdots b_n$$

因为

$$|A+B|=|P^H(A+B)P|=|P^HAP+P^HBP|$$

$$=\begin{vmatrix}a_1+b_1 & & & \\ & a_2+b_2 & & \\ & & \ddots & \\ & & & a_n+b_n\end{vmatrix}$$

$$=(a_1+b_1)(a_2+b_2)\cdots(a_n+b_n)$$

由 Hölder 不等式得

$$|A+B|^{\frac{1}{n}}=\{(a_1+b_1)\cdots(a_n+b_n)\}^{\frac{1}{n}}$$

$$\geqslant(a_1\cdots a_n)^{\frac{1}{n}}+(b_1\cdots b_n)^{\frac{1}{n}}$$

$$=|A|^{\frac{1}{n}}+|B|^{\frac{1}{n}}$$ 】

定理 5.4.6 设 A、B 为 n 阶正定 Hermite 矩阵，则

$$|A|+|B|<|A+B|$$

〔证明〕 由定理 5.4.5 知

$$|A+B|^{1/n}\geqslant|A|^{1/n}+|B|^{1/n}$$

由于 $|A|$、$|B|$ 均为正数，故

$$|A+B|\geqslant(|A|^{1/n}+|B|^{1/n})^n$$

$$=|A|+n|A|^{\frac{n-1}{n}}|B|^{\frac{1}{n}}+\cdots+|B|$$

$$>|A|+|B|$$ ▮

若用数学归纳法可把定理 5.4.5 推广为下述定理。

定理 5.4.7 设 A_1，A_2，…，A_k 是 k 个 n 阶正定 Hermite 矩阵，则有

$$|A_1+A_2+\cdots+A_k|^{1/n}\geqslant|A_1|^{1/n}+|A_2|^{1/n}+\cdots+|A_k|^{1/n}$$

且等号仅在 $A_i=C_iB(i=1, 2, \cdots, k)$ 时成立，C_i 为实数，B 为某矩阵。

〔证明〕 用数学归纳法。当 $k=2$ 时，由定理 5.4.5 知结论成立。

设不超过 $k-1$ 个矩阵时定理成立，下面讨论 k 个矩阵时的情况。

$$|A_1+A_2+\cdots+A_{k-1}+A_k|^{1/n}\geqslant|A_1+\cdots+A_{k-1}|^{1/n}+|A_k|^{1/n}$$

由归纳假设知

$$|A_1+\cdots+A_{k-1}|^{1/n}\geqslant|A_1|^{1/n}+\cdots+|A_{k-1}|^{1/n}$$

代入上式便得。其中等号仅在 $A_k=d(A_1+\cdots+A_{k-1})$，$A_i=C_iB(i=1, 2,\cdots,k-1)$ 时成立，因而 $A_i=C_iB(i=1, 2, \cdots, k-1, k)$ 成立。 ▮

§5.5 单纯矩阵偶在相似下的标准形

本节讨论两个单纯矩阵同时与对角矩阵相似的问题。为此先来给出关于两个可交换矩阵 A、$B(AB=BA)$ 的核、值域、不变子空间、特征子空间之间的关系。

定理 5.5.1 设 A、B 为两个 n 阶矩阵，且 $AB=BA$，则

(1) B 的核是 A 的不变子空间；反之，A 的核是 B 的不变子空间。

(2) B 的值域是 A 的不变子空间；反之，A 的值域是 B 的不变子空间。

(3) B 的任何特征子空间是 A 的不变子空间；反之，A 的任何特征子空间是 B 的不变子空间。

〔证明〕 (1) 设 B 的核为

$$N_B=\{\boldsymbol{x}\mid B\boldsymbol{x}=\boldsymbol{0}, \ \forall\boldsymbol{x}\in V\}$$

对于任何 $\boldsymbol{x}\in N_B$，因为 A、B 是可交换的，所以有

$$BA\boldsymbol{x}=AB\boldsymbol{x}=\boldsymbol{0}$$

这说明 $A\boldsymbol{x}\in N_B$，因此，N_B 是 A 的不变子空间。

（2） 设 B 的值域为

$$M_B=\{\boldsymbol{x}\mid\boldsymbol{x}=B\boldsymbol{u},\ \forall\boldsymbol{u}\in V\}$$

对于任何 $\boldsymbol{x}\in M_B$，有

$$A\boldsymbol{x}=AB\boldsymbol{u}=BA\boldsymbol{u}$$

因为 $A\boldsymbol{u}=\boldsymbol{v}\in V$，所以

$$A\boldsymbol{x}=B\boldsymbol{v}\in M_B$$

这说明 M_B 是 A 的不变子空间。

（3） 设 V_1 是 B 的特征值 λ 所对应的特征子空间，其维数为 r_1，λ 所对应的线性无关的特征向量为 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_{r_1}$（它恰构成 V_1 的一个基），任取 $\boldsymbol{u}\in V_1$，即

$$\boldsymbol{u}=\sum_{i=1}^{r_1}a_i\boldsymbol{x}_i,\ 且\ B\boldsymbol{u}=\lambda\boldsymbol{u}$$

由 A、B 可交换得

$$BA\boldsymbol{u}=\sum_{i=1}^{r_1}a_iAB\boldsymbol{x}_i=\sum_{i=1}^{r_1}a_i\lambda B\boldsymbol{x}_i=\lambda A\boldsymbol{u}$$

因此 $A\boldsymbol{u}\in V_1$，这表明 V_1 是 A 的不变子空间。 ▮

定理5.5.2 设 A、B 为两个 n 阶矩阵，且 $AB=BA$，则 B 的任何特征子空间中都有 A 的特征向量，反之，A 的任何特征子空间中都有 B 的特征向量。

〔证明〕 设 V_1 是 B 的特征值 λ 所对应的特征子空间，λ 所对应的线性无关的特征向量为 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_{r_1}$（它恰构成 V_1 的一个基）。由定理5.5.1知 V_1 是 A 的不变子空间，所以 $A\boldsymbol{x}_i\in V_1$，故设

$$A\boldsymbol{x}_i=\sum_{j=1}^{r_1}c_{ij}\boldsymbol{x}_j\qquad(i=1,\ 2,\ \cdots,\ r_1)\tag{5-49}$$

命

$$\boldsymbol{u}=\sum_{i=1}^{r_1}a_i\boldsymbol{x}_i\in V_1$$

欲使 $\boldsymbol{u}$ 是 A 的特征向量就应有

$$A\boldsymbol{u}=\mu\boldsymbol{u}$$

即

$$\sum_{i=1}^{r_1}a_iA\boldsymbol{x}_i=\mu\sum_{j=1}^{r_1}a_j\boldsymbol{x}_j$$

把式(5-49)代入上式得

$$\sum_{i=1}^{r_1}a_ic_{ij}\boldsymbol{x}_j=\mu\sum_{j=1}^{r_1}a_j\boldsymbol{x}_j$$

移项得

$$\sum_{j=1}^{r_1}\left(\sum_{i=1}^{r_1}a_ic_{ij}-\mu a_j\right)x_j=0$$

因为 x_1, x_2, …, x_{r_1}线性无关，所以

$$\sum_{i=1}^{r_1}a_ic_{ij}-\mu a_j=0 \qquad (j=1,\ 2,\ \cdots,\ r_1)$$

写成矩阵形式为

$$\begin{pmatrix} c_{11} & c_{21} & \cdots & c_{r_11} \\ c_{12} & c_{22} & \cdots & c_{r_12} \\ \cdots & \cdots & \cdots & \cdots \\ c_{1r_1} & c_{2r_1} & \cdots & c_{r_1r_1} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{r_1} \end{pmatrix}=\mu\begin{pmatrix} a_1 \\ \vdots \\ a_{r_1} \end{pmatrix}$$

式(5-50)表明$(a_1,\ a_2,\ \cdots,\ a_{r_1})^T$是矩阵

$$C=\begin{pmatrix} c_{11} & c_{21} & \cdots & c_{r_11} \\ c_{12} & c_{22} & \cdots & c_{r_12} \\ \cdots & \cdots & \cdots & \cdots \\ c_{1r_1} & c_{2r_1} & \cdots & c_{r_1r_1} \end{pmatrix}$$

的对应于特征值 μ 的特征向量。

综上所述，由与矩阵 C 的特征值 μ 相对应的特征向量$(a_1,\ a_2,\ \cdots,\ a_{r_1})^T$便能得到在 V_1 中属于 A 的特征向量。而对矩阵 C 来说，特征向量总是存在的。 ▌

推论 5.5.1 若 $AB=BA$，则 A 与 B 至少有一个公共的特征向量。

推论 5.5.2 若 $AB=BA$，λ_1，λ_2，…，λ_k 是 A 的 k 个相异特征值，则 A 与 B 至少有 k 个线性无关的公共特征向量。

推论 5.5.3 若 $AB=BA$，且 A 与 B 都有 n 个线性无关的特征向量，则 A 的任何特征子空间中都有 B 的个数与 A 相同的特征向量。

这三个推论的证明作为练习留给读者。

下面我们来讨论本节开始提出的问题。

定理 5.5.3 设 A、B 均为正规矩阵，则存在酉矩阵 U，使得

$$U^HAU=\text{diag}(a_1,\ a_2,\ \cdots,\ a_n)$$

与

$$U^HBU=\text{diag}(b_1,\ b_2,\ \cdots,\ b_n)$$

同时成立的充要条件是 $AB=BA$。

〔证明〕 必要性显然。现证充分性。

设 A 有 r 个相异特征值 λ_1，λ_2，…，λ_r，每一个 λ_i 的重数为 n_i，λ_i 对应的特征向量为 $x_1^{(i)}$，$x_2^{(i)}$，…，$x_{n_i}^{(i)}$，由于 A 的任何一个特征子空间都是 B 的不变子空间，所以

$$Bx_j^{(i)}=\sum_{k=1}^{n_i}c_{jk}^{(i)}x_k^{(i)} \qquad (j=1,\ 2,\ \cdots,\ n_i)$$

因为 A 是正规矩阵，可以假定所取的 $x_1^{(i)}$，…，$x_{n_i}^{(i)}$ 是标准正交向量组。命

$$X_i=(x_1^{(i)},\ x_2^{(i)},\ \cdots,\ x_{n_i}^{(i)})_{n\times n_i}$$

$$G_i=(c_{jk}^{(i)})_{n_i\times n_i}$$

不难验证下列二式是成立的

$$AX_i=\lambda_iX_i \qquad (i=1,\ 2,\ \cdots,\ r)$$

$$BX_i=X_iG_i$$

因为 X_i 的列向量组是标准正交向量组，故

$$X_i^HAX_i=\lambda_iI_{n_i}$$

$$X_i^HBX_i=G_i$$

现在证明 G_i 是正规的。根据定理 4.3.5 知，存在 $\tilde{X}_i$ 使得 n 阶矩阵 $(X_i,\ \tilde{X}_i)$ 是酉矩阵，命 $\tilde{G}_i=\tilde{X}_i^HB\tilde{X}_i$，则有

$$B(X_i,\tilde{X}_i)=(X_i,\tilde{X}_i)\begin{pmatrix}G_i & 0\\ 0 & \tilde{G}_i\end{pmatrix}$$

此即表明 B 与 $\mathrm{diag}(G_i,\ \tilde{G}_i)$ 酉相似，根据引理 5.1.5 知，$\mathrm{diag}(G_i,\ \tilde{G}_i)$ 是正规矩阵，从而 G_i 是 n_i 阶正规矩阵，它能与对角矩阵酉相似。即存在酉矩阵 F_i（n_i 阶），使得

$$G_iF_i=F_i\Lambda_i$$

其中 Λ_i 为 n_i 阶对角矩阵。

命

$$Y_i=X_iF_i \qquad (i=1,\ 2,\ \cdots,\ r)$$

易知，Y_i 的 n_i 个列向量是两两正交的单位列向量。因此，矩阵 $U=(Y_1,\ Y_2,\ \cdots,\ Y_r)$ 是 n 阶酉矩阵。于是

$$\begin{aligned}AU&=(AY_1,\ AY_2,\ \cdots,\ AY_r)\\&=(AX_1F_1,\ AX_2F_2,\ \cdots,\ AX_rF_r)\\&=(\lambda_1X_1F_1,\ \lambda_2X_2F_2,\ \cdots,\ \lambda_rX_rF_r)\\&=(\lambda_1Y_1,\ \lambda_2Y_2,\ \cdots,\ \lambda_rY_r)\\&=(Y_1,\ Y_2,\ \cdots,\ Y_r)\mathrm{diag}(\lambda_1I_{n_1},\ \cdots,\ \lambda_rI_{n_r})\end{aligned}$$

$$=U\mathrm{diag}(a_1, a_2, \cdots, a_n)$$

$$BU=(BY_1, BY_2, \cdots, BY_r)$$

$$=(BX_1F_1, BX_2F_2, \cdots, BX_rF_r)$$

$$=(X_1G_1F_1, X_2G_2F_2, \cdots, X_rG_rF_r)$$

$$=(X_1F_1\Lambda_1, X_2F_2\Lambda_2, \cdots, X_rF_r\Lambda_r)$$

$$=(Y_1\Lambda_1, Y_2\Lambda_2, \cdots, Y_r\Lambda_r)$$

$$=(Y_1, Y_2, \cdots, Y_r)\mathrm{diag}(\Lambda_1, \cdots, \Lambda_r)$$

$$=U\mathrm{diag}(b_1, b_2, \cdots, b_n)$$

因此，A、B 能同时与对角矩阵相似。 ▮

定理 5.5.4 设 n 阶矩阵 A、B 均能与对角矩阵相似，则存在 n 阶满秩矩阵 Y，使得

$$Y^{-1}AY=\mathrm{diag}(a_1, a_2, \cdots, a_n)$$

与

$$Y^{-1}BY=\mathrm{diag}(b_1, b_2, \cdots, b_n)$$

同时成立的充要条件是 $AB=BA$。

〔证明〕 必要性显然，现证充分性。

设A有 r 个相异特征值$\lambda_1, \lambda_2, \cdots, \lambda_r$，每一个 λ_i 的重数为 n_i，对应的特征向量为$x_1^{(i)}$，$x_2^{(i)}$，…，$x_{n_i}^{(i)}$，由于 A 的任何一个特征子空间都是 B 的不变子空间，所以

$$Bx_j^{(i)}=\sum_{k=1}^{n_i}c_{jk}^{(i)}x_k^{(i)} \qquad (j=1, 2, \cdots, n_i)$$

命

$$X_i=(x_1^{(i)}, x_2^{(i)}, \cdots, x_{n_i}^{(i)})_{n_i\times n_i}$$

$$G_i=(c_{jk}^{(i)})_{n_i\times n_i}$$

不难验证下列二式成立

$$AX_i=\lambda_iX_i$$

$$BX_i=X_iG_i$$

由推论 5.5.3 知 B 在 λ_i 特征子空间内恰有 n_i 个特征向量，由定理 5.5.1 的证明知道矩阵 G_i 有 n_i 个线性无关的特征向量，即 G_i 能与对角矩阵相似，所以存在 n_i 阶满秩矩阵 P_i，使得

$$G_iP_i=P_i\Lambda_i$$

其中 Λ_i 为 n_i 阶对角矩阵。若命

$$Y_i=X_iP_i \qquad (i=1, 2, \cdots, r)$$

容易证明n阶矩阵

$$Y=(Y_1,\ Y_2,\ \cdots,\ Y_r)$$

是满秩矩阵，且仿照定理5.5.3的证明可得

$$AY=Y\mathrm{diag}(a_1,\ a_2,\ \cdots,\ a_n)$$

$$BY=Y\mathrm{diag}(b_1,\ b_2,\ \cdots,\ b_n$$

习　题

5-1　设A是n阶矩阵，证明A^HA和AA^H有相同的特征值。

5-2　设A、B均为n阶正规矩阵，且$AB=BA$，则AB与BA均为n阶正规矩阵，且对于任何正整数k、l，A^kB^l是正规矩阵。

5-3　n阶复矩阵A是正规矩阵的充要条件是存在复矩阵A_1、A_2，使得

(1) $A=A_1+iA_2$

(2) $A_i=A_i^H$，$(i=1,\ 2)$

(3) $A_1A_2=A_2A_1$

5-4　n阶复矩阵A是正规矩阵的充要条件是存在正规矩阵B，使得$B^k=A$，其中k为任何正整数。

5-5　设A是n阶复矩阵，α、β均为复数，则$\alpha A+\beta A^H$是正规矩阵。

5-6　设A是n阶正规矩阵，且$A=A^2$，则$A=A^H$。

5-7　若$A=-A^H$，则$U=(A+I)(A-I)^{-1}$是酉矩阵。

5-8　设U是n阶酉矩阵，且$U-I$可逆，则$A=(U-I)^{-1}(U+I)$满足$A^H=-A$。

第六章 矩阵的分解

矩阵分解是矩阵理论及应用中常用的方法。将一个矩阵分解为比较简单的或性质比较熟悉的另一些矩阵的乘积或和，往往会使对这个被分解的矩阵的讨论和计算比较方便。

在前面各章的讨论中，对几种重要矩阵标准形的讨论实际上也是一种分解。

这章将介绍矩阵的正交三角分解（UR 或 QR 分解）、三角分解、奇异值分解、极分解和单纯矩阵的谱分解。

§6.1 矩阵的正交三角分解

矩阵正交三角分解简称 UR 分解或 QR 分解。在 §4.3 节中，曾运用Schmidt 正交化方法证明了满秩矩阵的 UR（或 QR）分解定理（定理 4.3.2 及推论 4.3.1）。这节将把正交三角分解定理推广到列满秩或行满秩的长方阵。

定义 6.1.1 设 A 为 $n\times l$ 复矩阵，如果 $\mathrm{rank}(A)=l$，则称 A 为列满秩，记为 $A\in C_l^{n\times l}$，如果 $\mathrm{rank}(A)=n$，则称 A 为行满秩，记为 $A\in C_n^{n\times l}$。

为了下文叙述方便起见，我们用 $U_l^{n\times l}$ 表示以 l 个两两正交的单位向量为列组成的矩阵的集合；以 $U_n^{n\times l}$ 表示以 n 个两两正交的单位向量为行组成的矩阵的集合。

定理 6.1.1 设 $A\in C_l^{n\times l}$，则存在 $Q\in U_l^{n\times l}$ 和满秩的 l 阶正线上三角复矩阵 R，使得

$$A=QR \tag{6-1}$$

并且分解式(6-1)是唯一的。

〔证明〕 因为 A 是列满秩的，所以对于任何 l 维列向量 $x\neq 0$，都有 $Ax\neq 0$。因此

$$x^HA^HAx=(Ax)^H(Ax)>0$$

这表示 A^HA 为 l 阶正定 Hermite 矩阵。于是存在唯一的 l 阶正线上三角矩阵 R，使得

$$A^HA==R^HR$$

令 $Q=AR^{-1}\in C^{n\times l}$，则

$$Q^HQ=R^{-H}A^HAR^{-1}=R^{-H}R^HRR^{-1}=I_l$$

上式表明 Q 的 l 个列向量是两两正交的单位向量，且有

$$A=QR$$

下面证明分解的唯一性。

若还有分解式 $A=Q_1R_1$，则

$$A^HA=R_1^HR_1=R^HR$$

因为正定 Hermite 矩阵 A^HA 的三角分解是唯一的，所以有

$$R=R_1$$

于是 $$Q=Q_1$$ ▮

定理 6.1.2 设 $A\in C_l^{n\times l}$，则存在 $Q\in U_n^{n\times l}$ 和满秩的 n 阶正线下三角复矩阵 L，使得

$$A=LQ \tag{6-2}$$

并且分解式(6-2)是唯一的。

〔证明〕 由定理 6.1.1 的证明知 AA^H 是 n 阶正定的 Hermite 矩阵，于是存在唯一的 n 阶正线下三角矩阵 L，使得

$$LL^H=A^HA$$

令 $Q=L^{-1}A\in C^{n\times l}$，则

$$QQ^H=I_n$$

于是，Q 的 n 个行向量是两两正交的单位向量。

唯一性的证明与定理 6.1.1 类似。 ▮

对于行满秩的长方阵，也有类似结论。

定理 6.1.3 设 $A\in C_n^{n\times l}$，则存在 $Q\in C_n^{n\times l}$ 和满秩的 n 阶正线下三角矩阵 L，使得

$$A=LQ \tag{6-3}$$

并且分解式(6-3)是唯一的；又存在 $Q_1\in U_l^{n\times l}$ 和满秩的 l 阶正线上三角矩阵 R，使得

$$A=Q_1R \tag{6-4}$$

并且分解式(6-4)是唯一的。

§6.2 矩阵的三角分解

在第三章中，已知正定的 Hermite 矩阵 A 可以分解成两个三角矩阵的乘积 $A=LL^H$，这节讨论一般矩阵的三角分解。

Gauss 消元法解方程组 $A\boldsymbol{x}=\boldsymbol{y}$ 的过程，用矩阵语言叙述就是把 A 分解成 $A=LR$，其中 L 为下三角矩阵，R 为上三角矩阵。一般说，这样的分解不是唯一的。例如，设 D 是行列式不为零的对角矩阵，则

$$A=LR=LD\,D^{-1}R=\tilde{L}\tilde{R}$$

$\tilde{L}$、$\tilde{R}$ 分别是下、上三角矩阵。

定理 6.2.1 设 $A\in C_n^{n\times n}$，则 A 可以唯一地分解为

$$A=\tilde{L}R \tag{6-5}$$

的充要条件是 A 的各阶顺序主子式

$$\Delta_k=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}\neq 0 \qquad (k=1,2,\cdots,n)$$

其中 L 为主对角线元素全是1的下三角矩阵（简称单位下三角矩阵），R 为上三角矩阵。

〔证明〕 设 $A=LR$，则

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}=\begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix}\begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}$$

其中 A_{11} 为 k 阶矩阵，A_{22} 为 $(n-k)$ 阶矩阵；L_{11}、L_{22} 分别为 k、$(n-k)$ 阶单位下三角矩阵；R_{11}、R_{22} 分别为 k、$(n-k)$ 阶单位上三角矩阵。于是有

$$A_{11}=L_{11}R_{11}$$

两边取行列式得

$$\det A_{11}=\Delta_k=\det L_{11}\det R_{11}=\det R_{11}$$

若设

$$R=\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix}$$

则有

$$\Delta_k=r_{11}r_{22}\cdots r_{kk} \qquad (k=1,2,\cdots,n-1)$$

$$\Delta_n=\det R=r_{11}r_{22}\cdots r_{nn}$$

因为 A 是满秩的，所以 $\Delta_n\neq 0$，故 r_{11}，r_{22}，…，r_{nn} 全不为零，因此

$$\Delta_k=r_{11}r_{22}\cdots r_{kk}\neq 0 \qquad (k=1,2,\cdots,n-1)$$

现用归纳法证明充分性。

当 A 的阶数为1时结论显然成立。设 A 的阶数为 $n-1$ 时有分解式(6-5)。现设 A 为 n 阶矩阵，则 A 可以写为

$$A=\begin{pmatrix} A_{n-1} & \beta \\ \alpha & a_{nn} \end{pmatrix}$$

因为

$$\Delta_k\neq 0 \qquad (k=1,2,\cdots,n-1)$$

故得 A_{n-1} 是 $n-1$ 阶满秩矩阵，即 A_{n-1}^{-1} 有意义，所以

$$\begin{pmatrix} I_{n-1} & 0 \\ -\alpha A_{n-1}^{-1} & 1 \end{pmatrix}A=\begin{pmatrix} I_{n-1} & 0 \\ -\alpha A_{n-1}^{-1} & 1 \end{pmatrix}\begin{pmatrix} A_{n-1} & \beta \\ \alpha & a_{nn} \end{pmatrix}$$

$$=\begin{pmatrix} A_{n-1} & \beta \\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta \end{pmatrix}$$

因此

$$A=\begin{pmatrix} I_{n-1} & 0 \\ -\alpha A_{n-1}^{-1} & 1 \end{pmatrix}^{-1}\begin{pmatrix} A_{n-1} & \beta \\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta \end{pmatrix}$$

$$=\begin{pmatrix} I_{n-1} & 0 \\ \alpha A_{n-1}^{-1} & 1 \end{pmatrix}\begin{pmatrix} A_{n-1} & \beta \\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta \end{pmatrix}$$

由归纳假设

$$A_{n-1}=\tilde{L}_1 R_1$$

得

$$A=\begin{pmatrix} I_{n-1} & 0 \\ \alpha A_{n-1}^{-1} & 1 \end{pmatrix}\begin{pmatrix} \tilde{L}_1 R_1 & \beta \\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta \end{pmatrix}$$

$$=\begin{pmatrix} I_{n-1} & 0 \\ \alpha A_{n-1}^{-1} & 1 \end{pmatrix}\begin{pmatrix} \tilde{L}_1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} R_1 & \tilde{L}_1^{-1}\beta \\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta \end{pmatrix}$$

$$=\begin{pmatrix} \tilde{L}_1 & 0 \\ \alpha A_{n-1}^{-1} L_1 & 1 \end{pmatrix}\begin{pmatrix} R_1 & \tilde{L}_1^{-1}\beta \\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta \end{pmatrix}$$

$$=\tilde{L}R$$

其中

$$\tilde{L}=\begin{pmatrix} \tilde{L}_1 & 0 \\ \alpha A_{n-1}^{-1}\tilde{L}_1 & 1 \end{pmatrix}$$

$$R=\begin{pmatrix} R_1 & \tilde{L}_1^{-1}\beta \\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta \end{pmatrix}$$

$\tilde{L}$ 与 R 分别为单位下三角矩阵与上三角矩阵。最后证明分解是唯一的。若

$$A=\tilde{L}R=\tilde{L}_1 R_1$$

则

$$\tilde{L}_1^{-1}\tilde{L}=R_1 R^{-1} \tag{6-6}$$

因为 $\tilde{L}_1^{-1}\tilde{L}$ 是单位下三角矩阵，而 R_1R^{-1} 是下三角矩阵，因此由式(6-6)得

$$\tilde{L}_1^{-1}L=R_1 R^{-1}=I$$

所以

$$\tilde{L}_1=\tilde{L},\qquad R_1=R$$

】

推论 6.2.1 设 $A\in C_n^{n\times n}$，则 A 可以唯一地分解为

$$A=L\tilde{R}$$

的充要条件是 A 的各阶顺序主子式

$$\Delta_k=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}\neq 0 \qquad (k=1,2,\cdots,n)$$

其中 L 为下三角矩阵，$\tilde{R}$ 为主对角线元素全是 1 的上三角矩阵（简称单位上三角矩阵）。

证明作为练习留给读者。

定理 6.2.2 设 $A\in C_n^{n\times n}$，则 A 可以唯一地分解为

$$A=\tilde{L}D\tilde{R} \tag{6-7}$$

的充要条件为 A 的各阶顺序主子式 $\Delta_k\neq 0\ (k=1,2,\cdots,n-1)$，其中 $\tilde{L}$、$\tilde{R}$ 分别是单位下、上三角矩阵，D 是对角矩阵。

〔证明〕 由定理 6.2.1 的证明知道，在 $A=\tilde{L}R$ 的表达式中，R 的主对角线元素全不为零，故可设

$$R=\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix}$$

$$=\begin{pmatrix} r_{11} & & & \\ & r_{22} & & \\ & & \ddots & \\ & & & r_{nn} \end{pmatrix}\begin{pmatrix} 1 & \dfrac{r_{12}}{r_{11}} & \cdots & \dfrac{r_{1n}}{r_{11}} \\ & 1 & \cdots & \dfrac{r_{2n}}{r_{22}} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}$$

$$=D\tilde{R}$$

其中 $r_{ii}=\Delta_i/\Delta_{i-1}\neq 0\ (i=1,2,\cdots,n)$，$D$ 为对角矩阵，$\tilde{R}$ 是单位上三角矩阵。由于 A 的三角分解式 (6-5) 是唯一的，所以式 (6-7) 也是唯一的。 ▌

式 (6-5) 称为 A 的 **LR 分解**，式 (6-7) 称为 A 的 **LDR 分解**

定理 6.2.3 设 A 为 n 阶矩阵，且 A 的所有前 $n-1$ 个顺序主子式均可逆，则 A 可以唯一地分解成

（1） $A=\tilde{L}R$，若 $\det A=0$，则 $r_{nn}=0$。

（2） $A=\tilde{L}D\tilde{R}$，若 $\det A=0$，则 $d_n=0$。

（3） $A=L\tilde{R}$，若 $\det A=0$，则 $l_{nn}=0$。

其中 $\tilde{L}=(\tilde{l}_{ij})$，$\tilde{R}=(\tilde{r}_{ij})$ 分别为单位下、上三角矩阵；$L=(l_{ij})$，$R=(r_{ij})$ 分别为下、上三角矩阵；$D=\mathrm{diag}(d_1,d_2,\cdots,d_n)$。

〔证明〕（1） 将 A 写成

$$A=\begin{pmatrix} A_{n-1} & \beta \\ \alpha & a_{nn} \end{pmatrix}$$

其中 A_{n-1} 是 $n-1$ 阶方阵，它的各阶顺序主子式均不为零，且 $\det A_{n-1}\neq 0$，由定理6.2.1知，A_{n-1} 可以唯一地分解成

$$A_{n-1}=\tilde{L}_1 R_1$$

式中 $\tilde{L}_1$ 是 $n-1$ 阶单位下三角矩阵，R_1 是 $n-1$ 阶上三角矩阵，$\tilde{L}_1$ 与 R_1 均可逆。

命

$$\tilde{L}=\begin{pmatrix}\tilde{L}_1 & 0\\ \alpha A_{n-1}^{-1}\tilde{L}_1 & 1\end{pmatrix}$$

$$R=\begin{pmatrix}R_1 & \tilde{L}_1^{-1}\beta\\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta\end{pmatrix}$$

那么不难验证 $\tilde{L}$ 是 n 阶单位下三角矩阵，R 是 n 阶上三角矩阵，且

$$A=\tilde{L}R$$

现证分解式的唯一性。若

$$A=\tilde{L}'R'$$

是 A 的另一种三角分解。将 $\tilde{L}'$ 与 R' 写成分块矩阵为

$$\tilde{L}'=\begin{pmatrix}\tilde{L}_2 & 0\\ \alpha A_{n-1}^{-1}\tilde{L}_2 & 1\end{pmatrix}$$

$$R'=\begin{pmatrix}R_2 & \tilde{L}_2^{-1}\beta\\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta\end{pmatrix}$$

于是

$$\begin{pmatrix}\tilde{L}_1 & 0\\ \alpha A_{n-1}^{-1}\tilde{L}_1 & 1\end{pmatrix}\begin{pmatrix}R_1 & \tilde{L}_1^{-1}\beta\\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta\end{pmatrix}$$

$$=\begin{pmatrix}\tilde{L}_2 & 0\\ \alpha A_{n-1}^{-1}\tilde{L}_2 & 1\end{pmatrix}\begin{pmatrix}R_2 & \tilde{L}_2^{-1}\beta\\ 0 & a_{nn}-\alpha A_{n-1}^{-1}\beta\end{pmatrix}$$

比较上式两端得

$$\tilde{L}_1R_1=\tilde{L}_2R_2$$

由于 $\tilde{L}_1$、$\tilde{L}_2$、R_1、R_2 均是可逆矩阵，所以

$$\tilde{L}_2^{-1}\tilde{L}_1=R_2R_1$$

上式左端是单位下三角矩阵，右端是上三角矩阵，因此

$$\tilde{L}_2^{-1}\tilde{L}_1=R_2R_1=I$$

即 $\qquad \tilde{L}_1=\tilde{L}_2,\qquad R_1=R_2$

于是 $\qquad \tilde{L}=\tilde{L}',\qquad R=R'$

最后，若 $\det A=0$，由于 $\tilde{L}_1$，R_1 均可逆，且 $\tilde{L}$ 为单位下三角矩阵，亦可逆，故得 $\det R=0$，于是 $r_{nn}=0$。

（2） 命 $d_i=r_{ii}$，则 $R=D\tilde{R}$。若 $\det A=0$，便有 $d_n=r_{nn}=0$。

（3） 与(1)的证明类似。 ■

定理 6.2.4 若$A=A^H$，且A的前$n-1$个顺序主子式均不为零，则A可以唯一地分解成

$$A=\tilde{L}D\tilde{L}^H \tag{6-8}$$

其中$D=\mathrm{diag}(d_1,d_2,\cdots,d_n)$，$d_i\in R(i=1,2,\cdots,n)$，$\tilde{L}$为单位下三角矩阵。

〔证明〕 由定理 6.2.3 知

$$A=\tilde{L}D\tilde{R}$$

其中$\tilde{L}$、$\tilde{R}$分别为单位下、上三角矩阵。由$A=A^H$得

$$A=\tilde{L}D\tilde{R}=\tilde{R}^HD^H\tilde{L}^H$$

因为$\tilde{L}$与$\tilde{R}^H$均是单位下三角矩阵，$\tilde{R}$与$\tilde{L}^H$均是单位上三角矩阵，所以根据矩阵LDR分解的唯一性得

$$\tilde{L}=\tilde{R}^H,\qquad D=D^H,\qquad d_i\in R$$

因此

$$A=\tilde{L}D\tilde{L}^H$$

】

推论 6.2.2 若$A=A^H$，且A是正定的，则A可以唯一地分解成

$$A=LL^H \tag{6-9}$$

其中L为正线下三角矩阵（参阅定理 4.6.7）。

〔证明〕 由定理 6.2.4 知

$$A=\tilde{L}D\tilde{L}^H$$

因为A的各阶顺序主子式

$$\Delta_k=d_{11}d_{22}\cdots d_{kk}>0\qquad(k=1,2,\cdots,n)$$

故令

$$D_1=\mathrm{diag}(\sqrt{d_1},\ \sqrt{d_2},\ \cdots,\ \sqrt{d_n})$$

$$L=\tilde{L}D_1$$

显然，L是正线下三角矩阵，且$D=D_1^2$，因此

$$A=\tilde{L}D\tilde{L}^H=LD_1^{-1}DD_1^{-H}L^H=LL^H$$

上述表达式的唯一性是显然的。 】

§6.3 矩阵的奇异值分解

矩阵的奇异值以及矩阵按奇异值的分解都是矩阵理论与应用中十分重要的内容。本节介绍矩阵的奇异值、性质以及矩阵按奇异值的分解。

命题 6.3.1 设$A\in C^{m\times n}$，则

$$\mathrm{rank}(AA^H)=\mathrm{rank}(A^HA)=\mathrm{rank}A \tag{6-10}$$

〔证明〕 若$\boldsymbol{x}$是$A^HA\boldsymbol{x}=\boldsymbol{0}$的非零解，则

$$x^H A^H A x = 0$$

于是 $$(Ax)^H(Ax) = 0$$

故得 $$Ax=0$$

这说明 x 也是 $Ax=0$ 的非零解。反之，若 x 是 $Ax=0$ 的非零解，易知它也是 $A^HAx=0$ 的非零解。因此方程组 $A^HAx=0$ 的系数矩阵的秩与 $Ax=0$ 的相等，即

$$\text{rank}A=\text{rank}A^HA$$

于是也有

$$\text{rank}A=\text{rank}A^H=\text{rank}\,AA^H$$ ▮

命题6.3.2 对于任何 $A\in C^{m\times n}$，A^HA 与 AA^H 均为半正定 Hermite 矩阵。

〔证明〕 由

$$(A^HA)^H=A^HA, \qquad (AA^H)^H=AA^H$$

知 A^HA、AA^H 均是 Hermite 矩阵。

因为对于任何 x 都有

$$(Ax)^H(Ax)\geqslant 0, \qquad (A^Hx)^H(A^Hx)\geqslant 0$$

所以 A^HA、AA^H 都是半正定的。 ▮

设 $A\in C_r^{m\times n}$，虽然 AA^H 与 A^HA 分别是 m 阶与 n 阶矩阵，但它们的秩都是 r，因此 AA^H 和 A^HA 的非零特征值都是 r 个，且全大于零，于是设

AA^H 的特征值 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_r>\lambda_{r+1}=\cdots=\lambda_m=0$

A^HA 的特征值 $\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_r>\mu_{r+1}=\cdots=\mu_n=0$

定理6.3.1 设 $A\in C_r^{m\times n}$，则

$$\lambda_i=\mu_i>0 \qquad (i=1,2,\cdots,r)$$

〔证明〕 设 x 是 AA^H 属于特征值 λ_i 的特征向量，即

$$AA^Hx=\lambda_i x$$

因此 $A^Hx\neq 0$，以 A^H 左乘上式两端得

$$A^HAA^Hx=\lambda_i A^Hx$$

这表明 A^Hx 是矩阵 A^HA 对应于特征值 λ_i 的特征向量，于是 AA^H 的全部特征值也必是 A^HA 的特征值。同理，A^HA 的全部特征值也必是 AA^H 的特征值。

设 x_1，x_2，…，x_p 是矩阵 AA^H 对应于特征值 $\lambda\neq 0$ 的 p 个线性无关的特征向量，由上面的讨论可知：A^Hx_1，A^Hx_2，…，A^Hx_p 是矩阵 A^HA 对应于 λ 的 p 个特征向量，现在证明 A^Hx_1，…，A^Hx_p 是线性无关的。事实上，

设 $$c_1A^Hx_1+c_2A^Hx_2+\cdots+c_pA^Hx_p=0$$

则 $$A(c_1A^Hx_1+\cdots+c_pA^Hx_p)=0$$

于是 $$c_1AA^Hx_1+\cdots+c_pAA^Hx_p=0$$

所以 $$\lambda(c_1x_1+\cdots+c_px_p)=0$$

由于 $\lambda\neq 0$，故

$$c_1x_1+c_2x_2+\cdots+c_px_p=0$$

因为 x_1，x_2，…，x_p 线性无关，所以

$$c_1=c_2=\cdots=c_p=0$$

此即 A^Hx_1，A^Hx_2，…，A^Hx_p 线性无关。

因为 AA^H 与 A^HA 的代数重复度等于几何重复度，所以当 $\lambda\neq 0$ 时，AA^H 的 p 重特征值也必是 A^HA 的 p 重特征值。类似地，A^HA 的 p 重特征值 μ 也必是 AA^H 的 p 重特征值。因此

$$\lambda_i=\mu_i \qquad (i=1,2,\cdots,r)$$ 】

根据定理6.3.1我们可以引进奇异值定义。

定义 6.3.1 设 $A\in C_r^{m\times n}$，AA^H 的正特征值为 λ_i，A^HA 的正特征值为 μ_i，称

$$\alpha_i=\sqrt{\lambda_i}=\sqrt{\mu_i} \qquad (i=1,2,\cdots,r) \tag{6-11}$$

是 A 的正奇异值，简称奇异值。

命题 6.3.3 A 与 A^H 有相同的正奇异值。

证明作为练习留给读者。

定义 6.3.2 设 A、$B\in C^{m\times n}$，如果存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V，使得

$$B=UAV \tag{6-12}$$

便说 A 与 B 是酉等价的。

定理 6.3.2 若 A 与 B 酉等价，则 A 与 B 有相同的正奇异值。

〔证明〕 因为

$$B=UAV$$

所以 $$B^HB=V^HA^HU^HUAV=V^HA^HAV$$

即 A^HA 与 B^HB 酉相似，故它们有相同的特征值，根据定义 6.3.1 知，A 与 B 有相同的正奇异值。 】

由第五章我们知道，一个正规矩阵酉相似于一个对角矩阵。而对于非正规矩阵我们有下面的定理

定理 6.3.3 设 $A\in C_r^{m\times n}$，$\alpha_1\geqslant\alpha_2\geqslant\cdots\geqslant\alpha_r$ 是 A 的 r 个正奇异值，则存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V，使得

$$U^HAV=D$$

其中 $$D=\begin{pmatrix}\Delta & 0\\ 0 & 0\end{pmatrix}$$

$\Delta=\mathrm{diag}(\delta_1,\delta_2,\cdots,\delta_r)$，$\delta_i$ 为复数，且 $|\delta_i|=\alpha_i$。

〔证明〕 因为 AA^H 为正规矩阵，所以存在 m 阶酉矩阵 U，使得

$$U^HAA^HU=\begin{pmatrix}\Delta\Delta^H & 0\\ 0 & 0\end{pmatrix}=\mathrm{diag}(\alpha_1^2,\ \cdots,\ \alpha_r^2,\ 0,\ \cdots,\ 0) \tag{6-13}$$

记 $$U=(x_1,\cdots,x_r,x_{r+1},\cdots,x_m)$$
$$=(U_1,\ U_2)$$

其中 $$U_1=(x_1,\cdots,x_r),\qquad U_2=(x_{r+1},\cdots,x_m)$$

代入式(6-13)得

$$(U_1,U_2)^HAA^H(U_1,U_2)$$
$$=\mathrm{diag}(\alpha_1^2,\cdots,\alpha_r^2,\ 0,\cdots,0)$$

比较上式两端，有

$$U_1^HAA^HU_1=\mathrm{diag}(\alpha_1^2,\cdots,\alpha_r^2)=\Delta^2=\Delta\Delta^H \tag{6-14}$$

$$U_2^HAA^HU_2=0 \tag{6-15}$$

由式(6-15)，有

$$(A^HU_2)^H(A^HU_2)=0$$

故 $$A^HU_2=0\quad 或\quad U_2^HA=0 \tag{6-16}$$

令 $$V_1\triangleq A^HU_1\Delta^{-H}\in C^{n\times r} \tag{6-17}$$

则 $$V_1^HV_1=\Delta^{-1}U_1^HAA^HU_1\Delta^{-H}$$

把式(6-14)代入上式，得

$$V_1^HV_1=\Delta^{-1}\Delta\Delta^H\Delta^{-H}=I_r$$

即 $V_1\in U^{n\times r}$，所以存在 $V_2\in U^{n\times(n-r)}$，使得

$$V=(V_1,V_2)\in U^{n\times n}$$

（参阅定理 4.3.5）。因此

$$0=V_1^HV_2=\Delta^{-1}U_1^HAV_2$$

或 $$U_1^HAV_2=0 \tag{6-18}$$

于是 $$U^HAV=\begin{pmatrix}U_1^H\\ U_2^H\end{pmatrix}A(V_1\ V_2)$$

$$=\begin{pmatrix} U_1^H A V_1 & U_1^H A V_2 \\ U_2^H A V_1 & U_2^H A V_2 \end{pmatrix}$$

根据式(6-16)、(6-17)与(6-18)得

$$U^H A V=\begin{pmatrix} U_1^H A A^H U_1 \Delta^{-H} & 0 \\ 0 & 0 \end{pmatrix}$$

再由式(6-14)便得

$$U^H A V=\begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}$$

】

推论 6.3.1 $m\times n$ 矩阵 A 与 B 酉等价的充要条件是 A 与 B 有相同的正奇异值。

证明作为练习留给读者。

定理 6.3.4 设 $A\in C_r^{m\times n}$，α_1，α_2，…，α_r 是 A 的正奇异值，正整数 l 满足 $r\leqslant l\leqslant \min(m,n)$，则总有 $U_l\in U^{m\times l}$，$V_l\in U^{n\times l}$，使得

$$U_l^H AV_l=\begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}\in C^{l\times l} \tag{6-19}$$

其中 $\Delta=\mathrm{diag}(\delta_1,\ \delta_2,\ \cdots,\ \delta_r)$，$\delta_i$ 为复数，且 $|\delta_i|=\alpha_i$，特别地，若 $l=r$，则

$$U_r^H A V_r=\Delta=\mathrm{diag}(\delta_1,\ \delta_2,\ \cdots,\ \delta_r)$$

〔证明〕 由定理 6.3.3 知，存在 $U\in U^{m\times m}$，$V\in U^{n\times n}$，使得

$$A=U D V^H$$

设 $$U=(u_1,u_2,\cdots,u_m),\qquad V=(v_1,v_2,\cdots,v_n)$$

则

$$A=\sum_{i=1}^{r}\delta_i u_i V_i^H$$

对于任何满足

$$r\leqslant l\leqslant \min(m,n)$$

的正整数 l，记

$$U_l=(u_1,u_2,\cdots,u_l)\in U^{m\times l}$$

$$V_l=(v_1,v_2,\cdots,v_l)\in U^{n\times l}$$

且命 $j>l$ 时 $\delta_j=0$，所以 $l\geqslant r$ 时有

$$A=\sum_{i=1}^{l}\delta_i u_i V_i^H=U_l D_l V_l^H \tag{6-20}$$

式中 $$D_l=\begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}\in C^{l\times l}$$

由式(6-20)得

$$U_l^H A V_l = D_l$$

当 $l=r$ 时，显然有

$$U_r^H A V_r = D_r = \Delta$$ ▌

推论 6.3.2 设 $A\in C_r^{m\times n}$，α_1，α_2，…，α_r 是 A 的正奇异值，则存在 $U\in U^{m\times m}$，$V\in U^{n\times n}$，使得

$$U^H A V=\begin{pmatrix}\Delta & 0\\ 0 & 0\end{pmatrix}\in R^{m\times n} \tag{6-21}$$

其中 $\Delta=\mathrm{diag}(\alpha_1,\alpha_2\cdots,\alpha_r)$，$\alpha_1\geqslant\alpha_2\geqslant\cdots\geqslant\alpha_r>0$。

式(6-21)称为矩阵奇异值分解，或称 UDV^H 分解。

推论 6.3.3 设 $A\in R_r^{m\times n}$，$\alpha_1\geqslant\alpha_2\geqslant\cdots\geqslant\alpha_r$ 是 A 的正奇异值，则存在正交矩阵 $P\in E^{m\times m}$，$Q\in E^{n\times n}$，使得

$$P^T A Q=\begin{pmatrix}\Delta & 0\\ 0 & 0\end{pmatrix}\in R^{m\times n} \tag{6-22}$$

其中 $\Delta=\mathrm{diag}(\alpha_1,\alpha_2\cdots,\alpha_r)$

作为推论 6.3.2 的一个推广，我们介绍如下定理。

定理 6.3.5 设 $A\in C_r^{m\times n}$，则存在 $P\in C_m^{m\times m}$，$Q\in C_n^{n\times n}$，使得

$$P A Q=\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix} \tag{6-23}$$

在关于矩阵的标准形的讨论中，这是一个主要结果。应用推论6.3.2很容易证明式(6-23)。根据式(6-23)可得

$$A=P^{-1}\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}Q^{-1}$$

$$=P^{-1}\begin{pmatrix}I_r\\ 0\end{pmatrix}_{m\times r}(I_r,0)_{r\times n}Q^{-1}$$

$$=BC^T \tag{6-24}$$

其中

$$B=P^{-1}\begin{pmatrix}I_r\\ 0\end{pmatrix}_{m\times r}\in C_r^{m\times r}$$

$$C^T=(I_r,0)_{r\times n}Q^{-1}\in C_r^{r\times n}$$

即 B、C 均为列满秩矩阵。式(6-24)常称为矩阵 A 的满秩分解式。

§6.4 矩阵的极分解

我们知道，任何一个非零的复数 z 总可以写成

$$z=\rho(\cos\alpha+\mathrm{i}\ \sin\alpha) \tag{6-25}$$

的形式，式中 $\rho>0$ 是 z 的模（或称极径），α 是 z 的幅角。把 z 写成这样的形式是唯一的，并称为复数的极分解。若把数看成是一阶矩阵，则 ρ 是一阶正定 Hermite 矩阵，$\cos\alpha+\mathrm{i}\ \sin\alpha$ 是一阶酉矩阵。我们称式(6-25)为一阶复矩阵的极分解。这节把极分解的观念推广到一般矩阵。

定理 6.4.1 设 $A\in C_n^{n\times n}$，则必存在酉矩阵 $U\in U^{n\times n}$ 与两个正定 Hermite 矩阵 H_1、H_2，使得

$$A=H_1U=UH_2 \tag{6-26}$$

且这样的分解式是唯一的。

分解式(6-26)称为矩阵 A 的极分解。

〔证明〕 因为 A 是满秩的，所以 A^HA 是正定 Hermite 矩阵，于是存在唯一的正定 Hermite 矩阵 H_2，使得

$$A^HA=H_2^2$$

从而 $$(AH_2^{-1})^H(AH_2^{-1})=H_2^{-H}A^HAH_2^{-1}=I$$

这表明 AH_2^{-1} 是 n 阶酉矩阵，命

$$AH_2^{-1}=U$$

故 $$A=UH_2$$

又 $$A=UH_2=UH_2U^HU$$

$$=(UH_2U^H)U=H_1U$$

其中 $H_1=UH_2U^H$，显然 H_1 是正定 Hermite 矩阵。

现证分解式是唯一的。设

$$A=H_1U=\tilde{H}_1\tilde{U}$$

故 $$H_1=\tilde{H}_1\tilde{U}U^H$$

从而 $$H_1^2=H_1H_1^H=\tilde{H}_1\tilde{U}U^HU\tilde{U}^H\tilde{H}_1^H=\tilde{H}_1^2$$

因此 $$H_1=\tilde{H}_1,\qquad U=\tilde{U}$$ ▮

推论 6.4.1 设 $A\in R_n^{n\times n}$，则存在正交矩阵 Q 与两个正定实对称矩阵 H_1、H_2，使得

$$A=H_1Q=QH_2$$

且分解式是唯一的。

推论 6.4.2 设 $A\in C_n^{n\times n}$，则存在酉矩阵 U_1、U_2，使得

$$U_2^HAU_1=\mathrm{diag}(\alpha_1,\alpha_2,\cdots,\alpha_n) \tag{6-27}$$

其中 $\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_n > 0$ 是 A 的 n 个正奇异值。

〔证明〕 由定理6.4.1知，存在酉矩阵 U 与正定矩阵 H，满足

$$A = UH, \qquad H^2 = A^H A \tag{6-28}$$

因为 H 是正定 Hermite 矩阵，所以存在酉矩阵 U_1，使得

$$H = U_1 \operatorname{diag}(\alpha_1, \cdots, \alpha_n) U_1^H \tag{6-29}$$

其中 α_1，$\cdots$，α_n 是 H 的 n 个特征值，即是 A 的 n 个正奇异值。

把式(6-29)代入式(6-28)得

$$A = U U_1 \operatorname{diag}(\alpha_1, \cdots, \alpha_n) U_1^H$$

命 $U_2 = U U_1$，则 U_2 是酉矩阵，代入上式得

$$A = U_2 \operatorname{diag}(\alpha_1, \cdots, \alpha_n) U_1^H$$

即

$$U_2^H A U_1 = \operatorname{diag}(\alpha_1, \cdots, \alpha_n)$$

∎

显然推论6.4.2是推论6.3.2的特殊情况。

推论6.4.3 设 $A \in R_n^{n \times n}$，则存在正交矩阵 Q_1 和 Q_2，使得

$$Q_2^T A Q_1 = \operatorname{diag}(\alpha_1, \cdots, \alpha_n) \tag{6-30}$$

其中 $\alpha_1, \cdots, \alpha_n$ 是 A 的 n 个正奇异值。

定理6.4.2 设 $A \in C^{n \times n}$，则存在酉矩阵 U 与两个半正定 Hermite 矩阵 H_1 与 H_2，使得

$$A = H_1 U = U H_2 \tag{6-31}$$

且 $H_1^2 = AA^H$，$H_2^2 = A^H A$。

〔证明〕 由推论6.3.2知，存在酉矩阵 U_1、U_2，使得

$$A = U_1 \begin{pmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{pmatrix} U_2$$

其中 α_1^2，$\alpha_2^2, \cdots, \alpha_n^2$ 是 A 的 n 个实特征值，$0 \leqslant \alpha_n \leqslant \cdots \leqslant \alpha_1$。因此

$$\begin{aligned} A &= \left(U_1 \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} U_1^H\right)(U_1 U_2) \\ &= (U_1 U_2)\left(U_2^H \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} U_2\right) \end{aligned} \tag{6-32}$$

若命

$$H_1 = U_1 \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} U_1^H$$

$$H_2=U_2^H\begin{pmatrix}\alpha_1 & & \\ & \ddots & \\ & & \alpha_n\end{pmatrix}U_2$$

$$U=U_1U_2$$

则
$$A=H_1U=UH_2$$

其中 H_1、H_2 是半正定 Hermite 矩阵，U 是酉矩阵。

根据式(6-38)可得

$$AA^H=U_1\begin{pmatrix}\alpha_1^2 & & \\ & \ddots & \\ & & \alpha_n^2\end{pmatrix}U_1^H=H_1^2$$

$$A^HA=U_2^H\begin{pmatrix}\alpha_1^2 & & \\ & \ddots & \\ & & \alpha_n^2\end{pmatrix}U_2=H_2^2$$

所以 H_1、H_2 唯一。 ▮

A 是实矩阵时，H_1、H_2 与 U 作相应的改变，读者自行叙述并证之。

§6.5 单纯矩阵的谱分解

由定理 1.7.1 知，n 阶单纯矩阵有 n 个线性无关的特征向量。

设 λ_1，λ_2，…，λ_n 是 A 的 n 个特征值，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_n$ 是 A 的 n 个线性无关的特征向量，且 $A\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i \quad (i=1,2,\cdots,n)$，命

$$P=(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_n) \tag{6-33}$$

$$\Lambda=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix} \tag{6-34}$$

则
$$A=P\Lambda P^{-1} \tag{6-35}$$

把式(6-35)两端取转置得

$$A^T=(P^T)^{-1}\Lambda P^T \tag{6-36}$$

这表明 A^T 也与对角矩阵相似。因此，设 $\boldsymbol{y}_1$，$\boldsymbol{y}_2$，…，$\boldsymbol{y}_n$ 是 A^T 的 n 个线性无关的特征向量，即

$$A^T\boldsymbol{y}_i=\lambda_i\boldsymbol{y}_i \qquad (i=1,2,\cdots,n) \tag{6-37}$$

把式(6-37)两端取转置得

$$\boldsymbol{y}_i^TA=\lambda_i\boldsymbol{y}_i^T \qquad (i=1,2,\cdots,n) \tag{6-38}$$

根据式(6-38)，我们称 $\boldsymbol{y}_i^T$ 是 A 的左特征向量，称 $\boldsymbol{x}_i$ 是 A 的右特征向量。根据式(6-36)知

$$(\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n)=(P^{\mathrm{T}})^{-1}=(P^{-1})^{\mathrm{T}} \tag{6-39}$$

把式(6-39)两端转置得

$$P^{-1}=\begin{pmatrix}\boldsymbol{y}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{y}_n^{\mathrm{T}}\end{pmatrix} \tag{6-40}$$

代入 $PP^{-1}=P^{-1}P=I$ 得

$$(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)\begin{pmatrix}\boldsymbol{y}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{y}_n^{\mathrm{T}}\end{pmatrix}=\begin{pmatrix}\boldsymbol{y}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{y}_n^{\mathrm{T}}\end{pmatrix}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)=I$$

此即

$$\boldsymbol{x}_1\boldsymbol{y}_1^{\mathrm{T}}+\boldsymbol{x}_2\boldsymbol{y}_2^{\mathrm{T}}+\cdots+\boldsymbol{x}_n\boldsymbol{y}_n^{\mathrm{T}}=I \tag{6-41}$$

$$\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{x}_j=\delta_{ij}\qquad(i,j=1,2,\cdots,n) \tag{6-42}$$

式(6-42)表明矩阵 A 的左特征向量与右特征向量正交。

把式(6-33)与式(6-40)代入式(6-35)得

$$A=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}\begin{pmatrix}\boldsymbol{y}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{y}_n^{\mathrm{T}}\end{pmatrix}$$

$$=\lambda_1\boldsymbol{x}_1\boldsymbol{y}_1^{\mathrm{T}}+\lambda_2\boldsymbol{x}_2\boldsymbol{y}_2^{\mathrm{T}}+\cdots+\lambda_n\boldsymbol{x}_n\boldsymbol{y}_n^{\mathrm{T}}$$

命

$$G_i=\boldsymbol{x}\boldsymbol{y}_i^{\mathrm{T}} \tag{6-43}$$

则

$$A=\sum_{i=1}^{n}\lambda_iG_i \tag{6-44}$$

式(6-44)称为单纯矩阵的谱分解。

设 A 有 r 个相异的特征值 λ_1，λ_2，…，λ_r，对应于 λ_j 的线性无关的右特征向量为 $\boldsymbol{x}_1^j$，$\boldsymbol{x}_2^j$，…，$\boldsymbol{x}_{\alpha_j}^j$，左特征向量为 $(\boldsymbol{y}_1^j)^{\mathrm{T}}$，$(\boldsymbol{y}_2^j)^{\mathrm{T}}$，…，$(\boldsymbol{y}_{\alpha_j}^j)^{\mathrm{T}}$，代入式(6-44)得

$$A=\sum_{j=1}^{r}\lambda_j\sum_{k=1}^{\alpha_j}\boldsymbol{x}_k^j(\boldsymbol{y}_k^j)^{\mathrm{T}}=\sum_{j=1}^{r}\lambda_jE_j \tag{6-45}$$

其中

$$E_j=\sum_{k=1}^{\alpha_j}\boldsymbol{x}_k^j(\boldsymbol{y}_k^j)^{\mathrm{T}}$$

$$=(\boldsymbol{x}_1^j,\ \boldsymbol{x}_2^j,\ \cdots,\ \boldsymbol{x}_{\alpha_j}^j)\begin{pmatrix}(\boldsymbol{y}_1^j)^{\mathrm{T}}\\ \vdots\\ (\boldsymbol{y}_{\alpha_j}^j)^{\mathrm{T}}\end{pmatrix} \tag{6-46}$$

根据式(6-41)得

$$\sum_{j=1}^{r} E_j = I$$

由式（6-42）得

$$E_j^2 = E_j$$

$$E_i E_j = 0 \qquad (i \neq j)$$

现在可以介绍谱分解的主要结论。

定理 6.5.1 设 A 是 n 阶单纯矩阵，λ_1，λ_2，…，λ_r 是 A 的 r 个相异特征值，则 A 可以进行满足下列性质的谱分解。

（1） $A = \sum_{j=1}^{r} \lambda_j E_j$

（2） $E_j^2 = E_j, \qquad (j = 1, 2, \cdots, r)$

（3） $E_i E_j = 0 \qquad (i \neq j,\ i,\ j = 1, 2, \cdots, r)$

（4） $\sum_{j=1}^{r} E_j = I$

（5） 满足上述四条性质的 $E_j (j = 1, 2, \cdots, r)$ 是唯一确定的。

（6） $\operatorname{rank} E_j = a_j \qquad (j = 1, 2, \cdots, r)$

〔证明〕（1）－（4）前面已证。现证（5）。若还有一个具有性质（1）-（4）但与 E_j 不同的 $\tilde{E}_j$ 存在，则

$$E_j A = E_j \left(\sum_{j=1}^{r} \lambda_j E_j \right) = \lambda_j E_j$$

$$= \left(\sum_{i=1}^{r} \lambda_i E_i \right) E_j = A E_j$$

类似地，

$$\tilde{E}_j A = A \tilde{E}_j = \lambda_j \tilde{E}_j$$

故得 $$E_j (A \tilde{E}_k) = E_j (\lambda_k \tilde{E}_k) = \lambda_k E_j \tilde{E}_k$$

$$(E_j A) \tilde{E}_k = \lambda_j E_j \tilde{E}_k$$

两式相减得

$$(\lambda_k - \lambda_j) E_j \tilde{E}_k = 0$$

因此当 $k \neq j$ 时

$$E_j \tilde{E}_k = 0$$

于是 $$E_j = E_j I = E_j \sum_{k=1}^{r} \tilde{E}_k = E_j \tilde{E}_j$$

$$=\left(\sum_{k=1}^{r} E_k\right)E_j=IE_j=E_j$$

性质(5)证毕。

最后证明性质(6)。由式(6-46)知，E_j 是 $n\times\alpha_j$ 矩阵与 $\alpha_j\times n$ 矩阵之积，所以

$$\operatorname{rank} E_j\leqslant\alpha_j \qquad (j=1,2,\cdots,r) \tag{6-47}$$

又由性质(4)知，

$$\operatorname{rank} I=n=\sum_{j=1}^{r}\alpha_j\leqslant\sum_{j=1}^{r}\operatorname{rank} E_j \tag{6-48}$$

综合式(6-47)与式(6-48)得

$$\alpha_j=\operatorname{rank} E_j \qquad (j=1,2,\cdots,r)$$

∎

第七章 范数 测度

我们知道，对于实数和复数，由于定义了它们的绝对值（模），可用模来表示其大小而带来许多方便。把这种单个数的模的概念推广到向量和矩阵，引入在某种意义上表示其"大小"的纯量，这就是向量范数（描述向量"大小"的量）和矩阵范数（描述矩阵"大小"的量）。向量范数和矩阵范数在矩阵分析中占有重要的地位。

本章讨论的内容是：向量范数，矩阵范数和矩阵测度。

§7.1 向量范数

先来讨论两个例子。

例 7.1.1 我们知道，平面向量 $\boldsymbol{x}=a\boldsymbol{i}+b\boldsymbol{j}$ 的模是用量$\sqrt{a^2+b^2}$来描述的。如果用 $\|\boldsymbol{x}\|$ 来表示 $\boldsymbol{x}$ 的模，那么 $\|\boldsymbol{x}\|$ 具有下面三条性质：

(1) 若 $\boldsymbol{x}\neq\boldsymbol{0}$，则 $\|\boldsymbol{x}\|>0$；当 $\boldsymbol{x}=\boldsymbol{0}$ 时，有 $\|\boldsymbol{x}\|=0$。

(2) $\|k\boldsymbol{x}\|=|k|\|\boldsymbol{x}\|$，$k$ 为任意实数。

(3) 对于任意平面向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$，有三角不等式

$$\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$$

例 7.1.2 由线性代数知，n 维欧氏空间中向量的模定义为

$$\|\boldsymbol{x}\|=(\boldsymbol{x},\boldsymbol{x})^{1/2}$$

这样定义的模也具有下面三条性质：

(1) 若 $\boldsymbol{x}\neq\boldsymbol{0}$，则 $\|\boldsymbol{x}\|>0$，$\|\boldsymbol{0}\|=0$。

(2) 对任意实数 k 和任意向量 $\boldsymbol{x}$，有

$$\|k\boldsymbol{x}\|=|k|\|\boldsymbol{x}\|$$

(3) 对任意向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$，有

$$\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$$

对于一般的线性空间，引入满足上述三条性质的纯量（或函数），用它来描述向量的大小，并称之为范数。

定义 7.1.1 设 F 为一数域，V 是 F 上的线性空间，对 $\forall\boldsymbol{x}\in V$，用 $\|\boldsymbol{x}\|$ 表示按照某个法则确定的与 $\boldsymbol{x}$ 对应的实数，且满足

(1) 非负性。当 $\boldsymbol{x}\neq\boldsymbol{0}$ 时，$\|\boldsymbol{x}\|>0$，对 $\forall\boldsymbol{x}\in V$，$\|\boldsymbol{0}\|=0$。

(2) 齐次性。$\|k\boldsymbol{x}\|=|k|\|\boldsymbol{x}\|$，对 $\forall\boldsymbol{x}\in V$，$k\in F$。

(3) 三角不等式。$\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$，对 $\forall\boldsymbol{x}$，$\boldsymbol{y}\in V$。

则称 $\|\boldsymbol{x}\|$ 为向量 $\boldsymbol{x}$ 的范数。

为了下面讨论的需要，先来证明两个重要的不等式。

Hölder 不等式 设$p>1$，$q=p/(p-1)$，则

$$\sum_{k=1}^{n}a_kb_k\leqslant\left(\sum_{k=1}^{n}a_k^p\right)^{1/p}\left(\sum_{k=1}^{n}b_k^q\right)^{1/q} \tag{7-1}$$

其中a_k，$b_k\geqslant0$。

〔证明〕 先来证明：如果u、v均非负，则总有

$$uv\leqslant\frac{u^p}{p}+\frac{v^q}{q} \tag{7-2}$$

考虑函数 $v=u^{p-1}$

不失一般性，考虑如图7-1所示的图形。

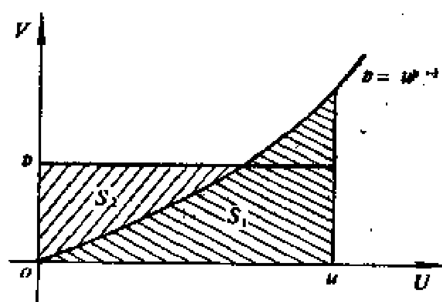

图 7-1

令 $$S_1=\int_0^u U^{p-1}\mathrm{d}U=\frac{1}{p}u^p$$

$$S_2=\int_0^v V^{\frac{1}{p-1}}\mathrm{d}V=\left(\frac{1}{p-1}+1\right)v^{\frac{1}{p-1}+1}$$

$$=\frac{1}{q}v^q$$

而$S_1+S_2\geqslant uv$，且等号只在$v=u^{p-1}$时成立。于是

$$uv\leqslant\frac{u^p}{p}+\frac{v^q}{q}$$

令 $$u=\frac{a_k}{a},\quad v=\frac{b_k}{b}$$

其中 $$a=\left(\sum_{k=1}^{n}a_k^p\right)^{\frac{1}{p}},\quad b=\left(\sum_{k=1}^{n}b_k^q\right)^{\frac{1}{q}}$$

代入式（7-2），得

$$\frac{a_k}{a}\cdot\frac{b_k}{b}\leqslant\frac{1}{p}\cdot\frac{a_k^p}{a^p}+\frac{1}{q}\cdot\frac{b_k^q}{b^q}$$

即 $$a_kb_k\leqslant ab\left(\frac{1}{p}\cdot\frac{a_k^p}{a^p}+\frac{1}{q}\cdot\frac{b_k^q}{b^q}\right)$$

于是 $$\sum_{k=1}^{n}a_kb_k\leqslant ab\left(\frac{1}{pa^p}\sum_{k=1}^{n}a_k^p+\frac{1}{qb^q}\sum_{k=1}^{n}b_k^q\right)$$

$$=ab\left(\frac{1}{p}+\frac{1}{q}\right)$$

$$=ab$$

$$=\left(\sum_{k=1}^{n}a_k^p\right)^{1/p}\left(\sum_{k=1}^{n}b_k^q\right)^{1/q}$$

令 $p=q=2$，代入式（7-1），有

$$\sum_{k=1}^{n} a_k b_k \leqslant \sqrt{\sum_{k=1}^{n} a_k^2}\ \sqrt{\sum_{k=1}^{n} b_k^2} \tag{7-3}$$

如果令 $u=\dfrac{f(x)}{\left(\int f(x)\mathrm{d}x\right)^{1/p}}$，$v=\dfrac{g(x)}{\left(\int g(x)\mathrm{d}x\right)^{1/q}}$

便得到积分形式的 Hölder 不等式：

$$\int f(x)g(x)\mathrm{d}x \leqslant \left(\int f(x)^p \mathrm{d}x\right)^{1/p}\left(\int g(x)^q \mathrm{d}x\right)^{1/q} \tag{7-4}$$

Minkowski 不等式 对任何 $p \geqslant 1$，有

$$\left(\sum_{i=1}^{n} |a_i+b_i|^p\right)^{1/p} \leqslant \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n} |b_i|^p\right)^{1/p} \tag{7-5}$$

〔证明〕 以 $q=p/(p-1)$ 代入下式

$$\sum_{i=1}^{n} |a_i+b_i|^p = \sum_{i=1}^{n} |a_i+b_i|\,|a_i+b_i|^{p-1}$$

有
$$\sum_{i=1}^{n} |a_i+b_i|^p = \sum_{i=1}^{n} |a_i+b_i|\,|a_i+b_i|^{\frac{p}{q}}$$

$$\leqslant \sum_{i=1}^{n} |a_i|\,|a_i+b_i|^{\frac{p}{q}} + \sum_{i=1}^{n} |b_i|\,|a_i+b_i|^{\frac{p}{q}}$$

由 Hölder 不等式，得

$$\sum_{i=1}^{n} |a_i+b_i|^p \leqslant \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |a_i+b_i|^p\right)^{\frac{1}{q}}$$

$$+\left(\sum_{i=1}^{n} |b_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |a_i+b_i|^p\right)^{\frac{1}{q}}$$

$$=\left[\left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |b_i|^p\right)^{\frac{1}{p}}\right]\left(\sum_{i=1}^{n} |a_i+b_i|^p\right)^{\frac{1}{q}}$$

于是有
$$\left(\sum_{i=1}^{n} |a_i+b_i|^p\right)^{\frac{1}{p}} \leqslant \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |b_i|^p\right)^{\frac{1}{p}}$$ ▮

由 Minkowski 不等式，我们可以引入常用的所谓 p- 范数的概念。

定义 7.1.2 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$，对任意 $p \geqslant 1$，称量

$$\left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \tag{7-6}$$

为向量 $\boldsymbol{x}$ 的 p- 范数，记作 $\|\boldsymbol{x}\|_p$。

易知，$\|\boldsymbol{x}\|_p$ 满足非负性和齐次性条件。由式（7-5）又有 $\|\boldsymbol{x}+\boldsymbol{y}\|_p \leqslant \|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p$。因此，$\|\boldsymbol{x}\|_p$ 满足范数的三条性质。

常用的 p- 范数有下面三种：

(1) $\|x\|_1$，即 $p=1$，这时

$$\|x\|_1=|x_1|+|x_2|+\cdots+|x_n|$$

(2) $\|x\|_2$，即 $p=2$，这就是例7.1.2中定义的欧氏空间中的向量范数，称为欧氏范数。

(3) $\|x\|_\infty$，即

$$\|x\|_\infty=\lim_{p\to}\|x\|_p$$

$$=\lim_{p\to\infty}=(|x_1|^p+\cdots+|x_n|^p)^{1/p}$$

命题 7.1.1 $\|x\|_\infty=\max_i|x_i|$

〔证明〕 令 $\alpha=\max_i|x_i|$，则

$$\left(\sum_{i=1}^n|x_i|^p\right)^{1/p}=\alpha\left(\sum_{i=1}^n\beta_i^p\right)^{1/p}$$

其中 $$\beta_i=\left|\frac{x_i}{\alpha}\right|\leqslant 1 \qquad (i=1,2,\cdots,n)$$

于是 $$\|x\|_p=\left(\sum_{i=1}^n|x_i|^p\right)^{1/p}=\left(\sum_{i=1}^n(\beta_i\alpha)^p\right)^{1/p}=\alpha\left(\sum_{i=1}^n\beta_i^p\right)^{1/p}$$

在 β_1，$\beta_2,\cdots,\beta_n$ 中至少有一个等于1，故

$$1\leqslant\left(\sum_{i=1}^n\beta_i^p\right)^{1/p}\leqslant n^{1/p}$$

因为 $\lim_{p\to\infty}n^{1/p}=1$，故

$$\lim_{p\to\infty}\left(\sum_{i=1}^n\beta_i^p\right)^{1/p}=1$$

因此 $$\|x\|_\infty=\lim_{p\to\infty}\|x\|_p=\alpha=\max_i|x_i|$$]

§7.2 向量范数的等价性

在一个有限维线性空间中，可以引进各种范数，按照不同法则规定的向量范数，其大小一般不等。例如，对向量 $x=(1,1,\cdots,1)$，有

$$\|x\|_2=\sqrt{n},\quad \|x\|_1=n,\quad \|x\|_\infty=1$$

虽然不同的范数可能有不同的量，但是这些范数之间仍有着重要的关系。譬如，在考虑向量序列收敛性问题时，它们就表现出了明显的一致性。这种性质称为范数的等价性。

定理 7.2.1 设 V 是 n 维线性空间，$\|x\|_a$ 和 $\|x\|_b$ 为任意两种向量范数（不限于 p- 范数），则总存在正数 c_1、c_2，对一切 $x\in V$，恒有

$$c_1\|\boldsymbol{x}\|_b \leqslant \|\boldsymbol{x}\|_a \leqslant c_2\|\boldsymbol{x}\|_b \qquad (7\text{-}7)$$

〔证明〕 先来证明形如式（7-7）的关系具有传递性。设任意两种范数 $\|\boldsymbol{x}\|_a$、$\|\boldsymbol{x}\|_b$ 都和一个固定的范数，例如 $\|\boldsymbol{x}\|_2$ 满足形如（7-7）的关系，即存在 c_1'、c_2' 及 c_1''、c_2''，使

$$c_1'\|\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{x}\|_a \leqslant c_2'\|\boldsymbol{x}\|_2$$

$$c_1''\|\boldsymbol{x}\|_b \leqslant \|\boldsymbol{x}\|_2 \leqslant c_2''\|\boldsymbol{x}\|_b$$

于是有
$$c_1'c_1''\|\boldsymbol{x}\|_b \leqslant \|\boldsymbol{x}\|_a \leqslant c_2'c_2''\|\boldsymbol{x}\|_b$$

这样，我们只须对 $b=2$ 证明关系式（7-7）即可。

设 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，…，$\boldsymbol{e}_n$ 是空间 V 的一组基底，则对 $\forall \boldsymbol{x}\in V$，有

$$\boldsymbol{x}=x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+\cdots+x_n\boldsymbol{e}_n$$

范数 $\|\boldsymbol{x}\|_a$ 是变量 x_1，x_2，…，x_n 的函数，记为

$$\|\boldsymbol{x}\|_a=\varphi(x_1,x_2,\cdots,x_n)$$

下面证明函数 $\varphi(x_1,x_2,\cdots,n_n)$ 是 x_1，$x_2,\cdots,x_n$ 的连续函数。设

$$\boldsymbol{x}'=x_1'\boldsymbol{e}_1+x_2'\boldsymbol{e}_2+\cdots+x_n'\boldsymbol{e}_n$$

$$\|\boldsymbol{x}'\|_a=\varphi(x_1',x_2',\cdots,x_n')$$

$$\begin{aligned}
&|\varphi(x_1',x_2',\cdots,x_n')-\varphi(x_1,x_2,\cdots,x_n)|\\
&\quad=|\,\|\boldsymbol{x}'\|_a-\|\boldsymbol{x}\|_a|\leqslant\|\boldsymbol{x}'-\boldsymbol{x}\|_a\\
&\quad=\|(x_1'-x_1)\boldsymbol{e}_1+(x_2'-x_2)\boldsymbol{e}_2+\cdots+(x_n'-x_n)\boldsymbol{e}_n\|_a\\
&\quad\leqslant|x_1'-x_1|\,\|\boldsymbol{e}_1\|_a+|x_2'-x_2|\,\|\boldsymbol{e}_2\|_a+\cdots+|x_n'-x_n|\,\|\boldsymbol{e}_n\|_a
\end{aligned}$$

因为 $\|\boldsymbol{e}_i\|_a$ 是常数，所以当 $|x_i'-x_i|$ 充分小时，有

$$|\varphi(x_1',x_2',\cdots,x_n')-\varphi(x_1,x_2,\cdots,x_n)|<\delta$$

这就证明了 $\varphi(x_1,x_2,\cdots,x_n)$ 是 $x_1,x_2,\cdots,x_n$ 的连续函数。

根据连续函数的性质，在有界闭集

$$x_1^2+x_2^2+\cdots+x_n^2=1 \qquad (7\text{-}8)$$

上，函数 $\varphi(x_1,x_2,\cdots,x_n)$ 可达最大值 M 和最小值 m。因为在式（7-8）中 x_i 不全为零，因此 $m>0$。记

$$d=\sqrt{\sum x_i^2}$$

则向量
$$\boldsymbol{y}=\frac{x_1}{d}\boldsymbol{e}_1+\frac{x_2}{d}\boldsymbol{e}_2+\cdots+\frac{x_n}{d}\boldsymbol{e}_n$$

的分量满足

$$\left(\frac{x_1}{d}\right)^2+\left(\frac{x_2}{d}\right)^2+\cdots+\left(\frac{x_n}{d}\right)^2=1$$

于是有

$$0<m\leqslant\|y\|_a=\varphi\left(\frac{x_1}{d},\frac{x_2}{d},\cdots,\frac{x_n}{d}\right)\leqslant M$$

但 $y=x/d$，故

$$md\leqslant\|x\|_a\leqslant Md$$

即
$$m\|x\|_2\leqslant\|x\|_a\leqslant M\|x\|_2$$
▌

定义 7.2.1 称关系式（7-7）为向量范数的等价关系，称满足式(7-7)的向量范数是等价的。

§7.3 矩阵范数

在这一节，我们进一步把范数的概念推广到 $m\times n$ 矩阵上。

定义 7.3.1 设 $A\in F^{m\times n}$（F 是实数域 R 或复数域 C）是任一个 $m\times n$ 矩阵，按照某个法则在 $F^{m\times n}$ 上规定 A 的一个实函数，记作 $\|A\|$，此函数具有下述性质：

(1) 非负性。若 $A\neq 0$，则 $\|A\|>0$，$\|0\|=0$。

(2) 齐次性。对任意纯量 k，$\|kA\|=|k|\,\|A\|$。

(3) 三角不等式。对任意 A，$B\in F^{m\times n}$，$\|A+B\|\leqslant\|A\|+\|B\|$。

则称 $\|A\|$ 为矩阵 A 的范数。

由于在矩阵理论和应用中，矩阵的乘法占有特殊重要的地位，矩阵和向量常以乘积的形式出现，因此，在讨论矩阵范数时，应考虑矩阵作用后的向量范数与矩阵范数及原向量范数的协调。为此，有下面的定义。

定义 7.3.2（范数相容性） 设在 $F^{m\times n}$，$F^{n\times p}$，$F^{m\times p}$ 定义了三种范数 $\|\cdot\|_{m\times n}$，$\|\cdot\|_{n\times p}$，$\|\cdot\|_{m\times p}$，对任意的矩阵 $A\in F^{m\times n}$，$B\in F^{n\times p}$，恒有

$$\|AB\|_{m\times p}\leqslant\|A\|_{m\times n}\|B\|_{n\times p} \tag{7-9}$$

则称范数 $\|\cdot\|_{m\times n}$，$\|\cdot\|_{n\times p}$ 和 $\|\cdot\|_{m\times p}$ 是相容的。

如果 $p=1$（这时 B 是 n 维列向量），式（7-9）就可写成

$$\|Ax\|_{m\times 1}\leqslant\|A\|_{m\times n}\|x\|_{n\times 1}$$

或写成

$$\|Ax\|_a\leqslant\|A\|_{m\times n}\|x\|_\beta$$

其中 $\|\cdot\|_a$ 和 $\|\cdot\|_\beta$ 表示 R^m 和 R^n 中的两种向量范数。

为了在矩阵分析的讨论中不致产生困难，我们所指的矩阵范数通常总是相容的。

下面我们来讨论几种具体的矩阵范数。

设 $x\in R^m$，$y\in R^n$，在相应的 m 维和 n 维线性空间中规定了向量的某种范数 $\|x\|_a$ 和 $\|y\|_\beta$。

设 $A\in F^{m\times n}$，规定矩阵 A 的范数如下：

$$\|A\| = \max_{\|x\|_\alpha = 1} \|Ax\|_\beta \qquad (7\text{-}10)$$

即矩阵 A 的范数取自当 $\|x\|_\alpha = 1$ 时，所有向量范数 $\|Ax\|_\beta$ 的最大值。由式（7-10）定义的矩阵范数称为诱导范数。

由式（7-10）定义的矩阵范数，满足定义 7.3.1 和相容性条件。

事实上，$\|0\| = 0$，且 $A \neq 0$ 时，$\|A\| > 0$，又

$$\|kA\| = \max_{\|x\|_\alpha = 1} \|kAx\|_\beta$$

$$= |k| \max_{\|x\|_\alpha = 1} \|Ax\|_\beta = |k| \|A\|$$

再有，对 $A, B \in F^{m\times n}$

$$\|A+B\| = \max_{\|x\|_\alpha = 1} \|(A+B)x\|_\beta$$

$$\leqslant \max_{\|x\|_\alpha = 1} \|Ax\|_\beta + \max_{\|x\|_\alpha = 1} \|Bx\|_\beta$$

$$= \|A\| + \|B\|$$

最后，显然有

$$\|Ax\|_\beta \leqslant \|A\| \|x\|_\alpha$$

即矩阵范数 $\|A\|$ 与向量范数 $\|x\|_\alpha$、$\|y\|_\beta$ 相容。

当 $\|x\|_\alpha$ 和 $\|y\|_\beta$ 分别取 $\|x\|_1$ 和 $\|y\|_1$，$\|x\|_2$ 和 $\|y\|_2$ 以及 $\|x\|_\infty$ 和 $\|y\|_\infty$ 时，我们得到三种矩阵范数，分别记为 $\|A\|_1$，$\|A\|_2$，$\|A\|_\infty$。

定理 7.3.1 设 $A \in F^{m\times n}$，则

(1) $\|A\|_1 = \max\limits_{j}\left(\sum\limits_{i=1}^{m} |a_{ij}|\right)$，$(j = 1, 2, \cdots, n)$

(2) $\|A\|_2 = \max\limits_{j}(\lambda_j(A^H A))^{1/2}$，$\lambda_j(A^H A)$ 表示矩阵 $A^H A$ 的第 j 个特征值。

(3) $\|A\|_\infty = \max\limits_{i}\left(\sum\limits_{j=1}^{n} |a_{ij}|\right)$，$(i = 1, 2, \cdots, m)$ $\qquad$ (7-11)

〔证明〕 (1) 令 $w = \max\limits_{j}\left(\sum\limits_{i=1}^{m} |a_{ij}|\right)$，$(j = 1, 2, \cdots, n)$。

设 $\|x\|_1 = 1$，$A = (\alpha_1, \alpha_2, \cdots, \alpha_n)$，则有

$$w = \max_j \|\alpha_j\|_1$$

又设 $x = (x_1, x_2, \cdots, x_n)^T$，于是

$$\|Ax\|_1 = \|x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n\|_1$$

$$\leqslant |x_1| \|\alpha_1\|_1 + |x_2| \|\alpha_2\|_1 + \cdots + |x_n| \|\alpha_n\|_1$$

$$\leqslant(|x_1|+|x_2|+\cdots+|x_n|)w$$

$$=\|\boldsymbol{x}\|_1 w=w$$

另一方面，设

$$\sum_{i=1}^{m}|a_{ir}|=w$$

取 $\boldsymbol{x}_r=(0,0,\cdots,\underset{r}{1},0,\cdots,0)$，则 $\|\boldsymbol{x}_r\|=1$，且

$$\|A\boldsymbol{x}_r\|_1=\sum_{i=1}^{m}|a_{ir}|=w$$

所以 $$\|A\|_1=\max\left(\sum_{i=1}^{m}|a_{ij}|\right),\qquad (j=1,2,\cdots,n)$$

(2) 若 $A\in F^{m\times n}$，$B\in F^{n\times p}$，因为对任意 $\boldsymbol{y}\in R^p$，有

$$\|AB\boldsymbol{y}\|_2\leqslant\|A\|_2\|B\boldsymbol{y}\|_2$$

$$\leqslant\|A\|_2\|B\|_2\|\boldsymbol{y}\|_2$$

故有 $$\|AB\|_2\leqslant\|A\|_2\|B\|_2$$

即矩阵范数 $\|\cdot\|_2$ 是相容的。

由向量范数 $\|\boldsymbol{x}\|_2$ 的定义知

$$\|A\boldsymbol{x}\|_2=(A\boldsymbol{x},\ A\boldsymbol{x})^{1/2}=(\boldsymbol{x},\ A^HA\boldsymbol{x})^{1/2}$$

因为 A^HA 是正定或半正定矩阵，所以 A^HA 的 n 个特征值非负，于是

$$\|A\|_2=\max_{\|\boldsymbol{x}\|_2=1}(A\boldsymbol{x},\ A\boldsymbol{x})^{1/2}$$

$$=\max_{\|\boldsymbol{x}\|_2=1}(\boldsymbol{x},\ A^HA\boldsymbol{x})^{1/2}$$

$$=\max_{\|\boldsymbol{x}\|_2=1}(\lambda_i(A^HA))^{1/2}$$

(3) 令 $w=\max\limits_{i}\left(\sum\limits_{j=1}^{n}|a_{ij}|\right)$，$(i=1,2,\cdots,m)$。对任意的 $\boldsymbol{x}\in R^n$，有

$$\|A\boldsymbol{x}\|_\infty=\max_{1\leqslant i\leqslant m}\left|\sum_{j=1}^{n}a_{ij}x_j\right|$$

$$\leqslant\max_{1\leqslant i\leqslant m}\left(\sum_{j=1}^{n}|a_{ij}|\right)\|\boldsymbol{x}\|_\infty=w\|\boldsymbol{x}\|_\infty$$

于是，若 $\|\boldsymbol{x}\|_\infty=1$，则有

$$\|A\|_\infty\leqslant w$$

另一方面，设

$$\sum_{j=1}^{n}|a_{ij}|=w$$

取 $x=(\text{sing}(a_{r1}),\ \text{sing}(a_{r2}),\ \cdots,\ \text{sing}(a_{rn}))^T$，则 $\|x\|_\infty=1$，且 $\|Ax\|_\infty=w$。]

范数 $\|A\|_1$ 称为列和范数，$\|A\|_\infty$ 称为行和范数，$\|A\|_2$ 称为谱范数。关于谱范数在下节将作进一步的讨论。

除了上面讨论的三种矩阵范数以外，还有一种常用的矩阵范数，就叫 Frobenius 范数。

设 $A\in F^{m\times n}$，由下式定义的范数

$$\|A\|_F=\left(\sum_{i=1}^{m}\ \sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} \tag{7-12}$$

叫做矩阵的 Frobenius 范数。

当 $A=x\in F^{n\times 1}$ 时，

$$\|A\|_F=\|x\|_2$$

所以 Frobenius 范数是向量范数 $\|x\|_2$ 的一个自然推广。范数 $\|A\|_F$ 实际上是把 A 看作 $F^{m\times n}$ 中的向量而建立的向量范数。

由（7-12）定义的矩阵范数显然满足定义 7.3.1 的条件。下面我们来证明范数 $\|A\|_F$ 满足相容性。

事实上，先考虑行矩阵 $A_1=(a_1,a_2,\cdots,a_n)$ 和分块矩阵 $B=(B_1,B_2,\cdots,B_e)^{n\times e}$

显然
$$\|A_1B\|_F^2=\sum_{j=1}^{e}|A_1B_j|^2$$

$$\leqslant\sum_{j=1}^{e}\|A_1^T\|_2^2\|B_j\|_2^2$$

$$=\|A_1\|_F^2\|B\|_F^2$$

再考虑
$$A=\begin{pmatrix}A_1\\A_2\\\vdots\\A_m\end{pmatrix}$$

有
$$\|AB\|_F^2=\sum_{i=1}^{m}\|A_iB\|_F^2\leqslant\sum_{i=1}^{m}\|A_i\|_F^2\|B\|_F^2$$

$$=\|A\|_F^2\|B\|_F^2$$

请读者自己验证，$\|A\|_F$ 与 $\|x\|_2$ 相容。

定理 7.3.2 设 $A\in C^{m\times n}$，U 和 V 分别为 m 阶和 n 阶酉矩阵，则

$$\|UAV\|_F=\|A\|_F$$

（证明） 令 $A=(A_1,A_2,\cdots,A_n)$，其中 A_i 为 A 的第 i 列，则

$$\|A\|_F^2=\|A_1\|_2^2+\|A_2\|_2^2+\cdots+\|A_n\|_2^2$$

由 $\|UA_i\|_2^2=\|A_i^HU^HUA_i\|_2=\|A_i\|_2^2$

有 $\|UA\|_F^2=\|A\|_F^2$

又 $\|A\|_F^2=\|A^H\|_F^2$

于是 $\|UAV\|_F^2=\|V^HA^HU^H\|_F^2$

$$=\|A^HU^H\|_F^2=\|UA\|_F^2=\|A\|_F^2$$]

最后我们指出，和向量范数一样，对于矩阵范数也有等价性定理，即：

设 $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 为两种矩阵范数，则总存在正数 $c_1>0$，$c_2>0$，对所有矩阵恒有

$$c_1\|A\|_b\leqslant\|A\|_a\leqslant c_2\|A\|_b$$

§7.4 矩阵的谱范数和谱半径

我们知道，矩阵范数 $\|A\|_2$ 称为谱范数，它是矩阵分析和系统理论中很有用的一种矩阵范数。从几何意义上来说，谱范数可以改写为

$$\|A\|_2=\max_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2}$$

它表示变换后向量的长 $\|Ax\|_2$ 与其原像长 $\|x\|$ 之比的最大值。

下面讨论谱范数的性质。

定理 7.4.1 设 $A\in C^{m\times n}$，则

(1) $\|A\|_2=\max\limits_{\|x\|=\|y\|=1}|y^HAx|$，$x\in F^n$，$y\in F^m$

(2) $\|A^H\|_2=\|A\|_2$

(3) $\|A^HA\|_2=\|A\|_2^2$

〔证明〕 (1) 对 $\|x\|=\|y\|=1$，有

$$|y^HAx|\leqslant\|y\|_2\|Ax\|_2\leqslant\|A\|_2$$

又设 $\|x\|=1$，并使 $\|Ax\|_2=\|A\|_2\neq 0$，于是若令 $y=Ax/\|Ax\|_2$，就有

$$|y^HAx|=\frac{\|Ax\|_2^2}{\|Ax\|_2}=\|Ax\|_2=\|A\|_2$$

(2) $$\|A\|_2=\max_{\|x\|=\|y\|=1}|y^HAx|$$

$$=\max_{\|x\|=\|y\|=1}|x^HA^Hy|=\|A^H\|_2$$

(3) 由 $\|A^HA\|_2\leqslant\|A^H\|_2\|A\|_2$，$\|A^H\|_2=\|A\|_2$，知

$$\|A^HA\|_2\leqslant\|A\|_2^2 \tag{7-13}$$

令 $\|x\|_2=1$，$\|Ax\|_2=\|A\|_2$，于是

$$\|A^H A\|_2 \geq \max_{\|x\|=1} |x^H A^H A x|$$

$$= \max_{\|x\|=1} \|Ax\|_2^2 = \|A\|_2^2 \quad (7\text{-}14)$$

由式（7-13）和式（7-14）即知 (3) 成立。 ∎

定理 7.4.2 设 $A \in C^{m\times n}$，$U \in C^{m\times m}$，$V \in C^{n\times n}$，且 $U^H U = I_m$，$V^H V = I_n$，则

$$\|UAV\|_2 = \|A\|_2 \quad (7\text{-}15)$$

〔证明〕 令 $v = Vx$，$u = U^H y$，则

$\|x\|_2 = 1$ 当且仅当 $\|v\|_2 = 1$

$\|y\|_2 = 1$ 当且仅当 $\|u\|_2 = 1$

于是

$$\|A\|_2 = \max_{\|x\|_2 = \|y\|_2 = 1} |y^H A x|$$

$$= \max_{\|v\|_2 = \|u\|_2 = 1} |u^H UAV v|$$

$$= \|UAV\|_2$$ ∎

定理 7.4.3 若 $\|A\|_2 < 1$，则 $I - A$ 为非奇异，且

$$\|(I-A)^{-1}\|_2 \leq (1 - \|A\|_2)^{-1} \quad (7\text{-}16)$$

〔证明〕 设 x 为任一非零向量，则

$$\|(I-A)x\|_2 = \|x - Ax\|_2$$

$$\geq \|x\|_2 - \|Ax\|_2$$

$$\geq \|x\|_2 - \|A\|_2 \|x\|_2$$

$$= (1 - \|A\|_2)\|x\|_2 > 0$$

于是，若 $x \neq 0$，则 $(I-A)x \neq 0$，从而方程

$$(I-A)x = 0$$

无非零解，故矩阵 $I-A$ 非奇异。

因为 $I-A$ 非奇异，故有

$$(I-A)(I-A)^{-1} = I$$

于是

$$(I-A)^{-1} = [(I-A) + A](I-A)^{-1}$$

$$= (I-A)(I-A)^{-1} + A(I-A)^{-1}$$

$$= I + A(I-A)^{-1}$$

从而

$$\|(I-A)^{-1}\|_2 \leq \|I\|_2 + \|A\|_2 \|(I-A)^{-1}\|_2$$

$$= 1 + \|A\|_2 \|(I-A)^{-1}\|_2$$

即 $$\|(I-A)^{-1}\|_2\leqslant\frac{1}{1-\|A\|_2}$$ ▮

下面我们引入一个在矩阵分析中经常用到的概念——**谱分解**。

定义 7.4.4 设 $A\in F^{n\times n}$，$\lambda_1,\lambda_2,\cdots,\lambda_n$ 为A的特征值，称

$$\rho(A)=\max_i(|\lambda_i|,\ i=1,2,\cdots,n)$$

为 A 的**谱半径**。

谱半径 $\rho(A)$ 在几何上表示以原点为圆心、能包含 A 的全部特征值的圆的半径的最小值。

定理 7.4.2 对任意矩阵 $A\subset F^{n\times n}$，总有

$$\rho(A)\leqslant\|A\|_2 \tag{7-17}$$

〔证明〕 设 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 为 A 的特征值，不妨设 $|\lambda_k|=\rho(A)$。令 $\boldsymbol{x}_k$ 为 A 对应于 λ_k 的特征向量，则有

$$\|A\boldsymbol{x}_k\|_2=|\lambda_k|\|\boldsymbol{x}_k\|_2=\rho(A)\|\boldsymbol{x}_k\|_2$$

于是 $$\rho(A)\leqslant\|A\|_2$$ ▮

特别地，如果 A 为正规矩阵 $(AA^H=A^HA)$，则有下面的推论成立。

推论 7.4.1 设 $A\in F^{n\times n}$，且 A 为正规矩阵，则

$$\|A\|_2=\rho(A) \tag{7-18}$$

〔证明〕 由于 A 是正规的，故有U，$U^HU=I$，使

$$U^HAU=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

于是 $$\|A\|_2=\|\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\|_2$$

$$=\max\left[\sum_{i=1}^n|\lambda_i\xi_i|^2;\quad\sum_{i=1}^n\xi_i^2=1\right]^{\frac{1}{2}}$$

$$\leqslant\rho(A)$$

由式(7-17)，有

$$\rho(A)\leqslant\|A\|_2$$

于是 $$\rho(A)=\|A\|_2$$ ▮

§7.5 矩阵测度

矩阵测度是和矩阵范数有密切关系的一个量，它也是矩阵分析和系统理论中经常用到的概念。

定义 7.5.1 设矩阵 $A\in F^{n\times n}$，称由下式确定的纯量

$$\lim_{h\to 0^+}\frac{\|I_n+hA\|-1}{h} \tag{7-19}$$

为矩阵 A 的测度，记作 $\mu(A)$，其中 $\|I_n+hA\|$ 是由某向量范数诱导的矩阵范数。

因为 $\|I\|=1$，所以式（7-19）意味着矩阵范数 $\|\cdot\|$ 在点 I 上沿 A 的方向导数。

例 7.5.1 对于向量范数 $\|x\|_1 \triangleq \sum_{i=1}^{n}|x_i|$

$$\|A\|_1=\max_j\sum_{i=1}^{n}|a_{ij}|$$

有

$$\mu_1(A)=\max_j\left\{\mathrm{Re}(a_{jj})+\sum_{\substack{i=1\\i\neq j}}^{n}|a_{ij}|\right\} \tag{7-20}$$

这是因为

$$\|I+hA\|=\max_j\sum_{i=1}^{n}|\delta_{ij}+ha_{ij}|$$

$$=\max_j\left\{|1+ha_{jj}|+\sum_{\substack{i=1\\i\neq j}}^{n}|ha_{ij}|\right\}$$

其中 $i=j$ 时 $\delta_{ij}=1$，$i\neq j$ 时 $\delta_{ij}=0$。

对于足够小的 $h>0$，有

$$\|I+hA\|_1=\max_j\left\{1+h\mathrm{Re}(a_{jj})+O(h)+\sum_{\substack{i=1\\i\neq j}}^{n}|ha_{ij}|\right\}$$

其中 $O(h)$ 在 $h\to 0^+$ 时 $O(h)/h\to 0$ 。

于是

$$\mu_1(A)=\lim_{h\to 0^+}\frac{\|I_n+hA\|-1}{h}$$

$$=\max_j\left\{\mathrm{Re}(a_{jj})+\sum_{\substack{i=1\\i\neq j}}^{n}|a_{ij}|\right\}$$

同样，对于向量范数 $\|x\|_\infty=\max_i|x_i|$，有

$$\|A\|_\infty=\max_i\sum_{j=1}^{n}|a_{ij}|$$

$$\mu_\infty(A)=\max_i\left\{\mathrm{Re}(a_{ii})+\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|\right\} \tag{7-21}$$

下面我们来讨论矩阵测度的性质。

命题 7.5.1 $\mu(0)=0$，$\mu(I_n)=1$，$\mu(-I_n)=-1$ 。

〔证明〕 $\mu(0)=0$ 是显然的，但需要指出，$\mu(A)=0$ 并不意味着 $A=0$ 。

现在来证明后两个等式。

因为矩阵范数 $\|I_n+hA\|$ 为某向量范数的诱导范数，$\|I_n\|=1$，于是

$$\mu(I_n)=\lim_{h\to 0^+}\frac{\|I_n+hI_n\|-1}{h}$$

$$=\lim_{h\to 0^+}\frac{(1+h)\|I_n\|-1}{h}=1$$

$$\mu(-I_n)=\lim_{h\to 0^+}\frac{\|I_n-hI_n\|-1}{h}$$

$$=\lim_{h\to 0^+}\frac{(1-h)\|I_n\|-1}{h}=-1$$ ▌

命题7.5.1表明，由于A的形式不同，$\mu(A)$可能取负值。

命题7.5.2 $\mu(kA)=k\mu(A)$，$\mu(A+kI)=\mu(A)+k$，其中k为非负实数，$\forall A\in F^{n\times n}$。

〔证明〕 按定义7.5.1，有

$$\mu(kA)=\lim_{h\to 0^+}\frac{\|I+h(kA)\|-1}{h}$$

$$=\lim_{h\to 0^+}k\frac{\|I+hkA\|-1}{hk}=k\mu(A)$$

$$\mu(A+kI)=\lim_{h\to 0^+}\frac{\|I+hA+hkI\|-1}{h}$$

$$=\lim_{h\to 0^+}\frac{(1+hk)\|I+\frac{h}{1+hk}A\|-1}{h}$$

$$=\lim_{h\to 0^+}\frac{\|I+\frac{h}{1+hk}A\|+hk\|I+\frac{h}{1+hk}A\|-1}{h}$$

$$=\lim_{h\to 0^+}\frac{\|I+\frac{h}{1+hk}A\|-1}{h}+\lim_{h\to 0^+}k\|I+\frac{h}{1+hk}A\|$$

$$=\mu(A)+k\|I\|=\mu(A)+k$$ ▌

命题7.5.3 $-\|A\|\leqslant-\mu(A)\leqslant\mu(A)\leqslant\|A\|$，$\forall A\in F^{n\times n}$

〔证明〕 按范数定义的三角不等式，有

$$\frac{\|I-hA\|}{h}+\frac{\|I+hA\|}{h}\geqslant\frac{\|2I\|-2}{h}=\frac{2-2}{h}=0$$

又

$$\frac{\|I+hA\|}{h}\leqslant\frac{\|I\|+h\|A\|-1}{h}=\|A\|$$

$$-\|A\|=\frac{-h\|A\|+1-1}{h}\leqslant-\frac{\|I-hA\|-1}{h}$$

由上面三个不等式，得

$$-\|A\|\leqslant-\frac{\|I-hA\|-1}{h}\leqslant\frac{\|I+hA\|-1}{h}\leqslant\|A\|$$

当 $h\to0^+$ 时，有

$$-\|A\|\leqslant-\mu(A)\leqslant\mu(A)\leqslant\|A\|$$ 】

命题7.5.3表明，在范数和测度之间存在大小关系：$|\mu(A)|\leqslant\|A\|$。

命题7.5.4 $\mu(A+B)\leqslant\mu(A)+\mu(B)$，$\forall A\in F^{n\times n}$，$B\in F^{n\times n}$。

〔证明〕 因为

$$\frac{\|I+h(A+B)\|-1}{h}=\frac{\|(I+2hA)+(I+2hB)\|-2}{2h}$$

$$\leqslant\frac{\|I+2hA\|-1}{2h}+\frac{\|I+2hB\|-1}{2h}$$

故有 $$\mu(A+B)\leqslant\mu(A)+\mu(B)$$ 】

命题7.5.5 设 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 为 A 的特征值，则

$$-\mu(-A)\leqslant\mathrm{Re}\lambda_i(A)\leqslant\mu(A)\qquad(i=1,2,\cdots,n)$$

〔证明〕 设矩阵 A 对应于特征值 λ_i 的特征向量为 x，并且 $\|x\|=1$，则按矩阵范数定义，有

$$\frac{\|I+hA\|\cdot\|x\|-1}{h}\geqslant\frac{\|x+hAx\|-1}{h}$$

$$=\frac{\|x+h\lambda_i x\|-1}{h}=\frac{|1+h\lambda_i|\|x\|-1}{h}$$

$$=\frac{|1+h\lambda_i|-1}{h}=\frac{h\mathrm{Re}(\lambda_i)+O(h)}{h}$$

于是 $$\mu(A)\geqslant\mathrm{Re}(\lambda_i)\tag{7-22}$$

再由 $$\frac{\|I-hA\|\cdot\|x\|-1}{h}\geqslant\frac{\|x-hAx\|-1}{h}$$

$$=\frac{|1-h\lambda_i|-1}{h}=\frac{-h\mathrm{Re}(\lambda_i)+O(h)}{h}$$

有 $$-\frac{\|I-hA\|\cdot\|x\|-1}{h}\leqslant\mathrm{Re}(\lambda_i)-\frac{O(h)}{h}$$

于是 $$-\mu(-A)\leqslant\mathrm{Re}(\lambda_i)\tag{7-23}$$

由式（7-22）和式（7-23），有

$$-\mu(-A)\leqslant\mathrm{Re}\lambda_i(A)\leqslant\mu(A)$$ 】

命题 7.5.6 对任意向量 $\boldsymbol{x}\in F^n$ 和矩阵 $A\in F^{n\times n}$，有

$$\|A\boldsymbol{x}\|\geqslant\max\{-\mu(-A),\ -\mu(A)\}\|\boldsymbol{x}\| \qquad (7\text{-}24)$$

〔证明〕

$$\|A\boldsymbol{x}\|=\frac{\|(I+hA)\boldsymbol{x}-\boldsymbol{x}\|}{h}\geqslant-\frac{\|I+hA\|-1}{h}\|\boldsymbol{x}\|$$

于是 $$\|A\boldsymbol{x}\|\geqslant-\mu(A)\|\boldsymbol{x}\|$$

又 $$\|A\boldsymbol{x}\|=\frac{\|(I-hA)\boldsymbol{x}-\boldsymbol{x}\|}{h}\geqslant\frac{1-\|I-hA\|}{h}\|\boldsymbol{x}\|$$

于是 $$\|A\boldsymbol{x}\|\geqslant-\mu(-A)\|\boldsymbol{x}\|$$

故得 $$\|A\boldsymbol{x}\|\geqslant\max\{-\mu(-A),\ -\mu(A)\}\|\boldsymbol{x}\|$$ ▌

定理 7.5.1 对于由任意 $A\in C^{n\times n}$ 和任意向量范数导出的矩阵范数，$\mu(A)$ 存在。

〔证明〕 对于 $0<\varepsilon<1$ 的实数 ε，有

$$\frac{\|I+\varepsilon hA\|-1}{\varepsilon h}=\frac{\|\varepsilon(I+hA)+(1-\varepsilon)I\|-1}{\varepsilon h}\leqslant\frac{\|I+hA\|-1}{h}$$

因此，函数 $f(h)\triangleq\{\|I+hA\|-1\}/h$ 在 $h>0$ 处连续，在 $h\to0^+$ 时减小，而且 $f(h)\geqslant-\|A\|$（命题 7.5.3），因此，$\lim\limits_{h\to0} f(h)=\mu(A)$。 ▌

习　题

7-1 证明

(1) $$\|\boldsymbol{x}-\boldsymbol{y}\|\geqslant\|\boldsymbol{x}\|-\|\boldsymbol{y}\|$$

(2) $$\|A-B\|\geqslant\|A\|-\|B\|$$

7-2 设 $A=(a_{ij})_{n\times n}$ 是正定对称矩阵，证明：不论 $x_1,x_2,\cdots,x_n$ 和 $y_1,y_2,\cdots,y_n$ 是何实数，恒有不等式

$$(\sum a_{ij}x_ix_j)^2\leqslant(\sum a_{ij}x_ix_j)(\sum a_{ij}y_iy_j)$$

7-3 设 $A=(a_{ij})_{n\times n}$，$\|A\|$ 是向量范数诱导的矩阵范数。证明 $\|A\|\geqslant|a_{ij}|$。

7-4 设 $\|A\|$ 是诱导范数，$\det A\neq0$。证明：

(1) $$\|A^{-1}\|\geqslant\|A\|^{-1}$$

(2) $$\|A^{-1}\|^{-1}=\min_{\boldsymbol{x}\neq0}\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$$

7-5 证明下列不等式

(1) $$\max\{\mu(A)-\mu(-B),\ -\mu(A)+\mu(B)\}\leqslant\mu(A+B)$$

(2) $$|\mu(A)-\mu(B)|\leqslant\max\{|\mu(A-B)|,\ |\mu(B-A)|\}$$

7-6　证明：　$\mu(A-B)\geqslant\mu(A)-\mu(B)$。

7·7　证明下列不等式

（1）　$$\|A^{-1}\|^{-1}\geqslant\max\{-\mu(-A),\ -\mu(A)\},\ \det A\neq 0$$

（2）　$$\|(I+hA)^{-1}\|^{-1}\geqslant 1-h\mu(-A)\geqslant 1-h\|A\|$$

7·8　证明：

$$\mu_{\infty}(A)=\max_{i}\left\{\operatorname{Re}(a_{ii})+\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\right\}$$

第八章　矩阵序列和矩阵级数

上一章我们引进了向量范数和矩阵范数这样的纯量，这一章我们将利用范数来讨论向量序列和矩阵序列的极限和矩阵级数。

§8.1　向量序列与极限

定义 8.1.1　将域F上向量空间$V(F)$的范数$\|\cdot\|$固定①。$V(F)$的向量序列$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_m$，…，用$\{\boldsymbol{x}_m\}$表示，对于$\{\boldsymbol{x}_m\}$，当存在某个向量$\boldsymbol{x}\in V(F)$，使得$m\to\infty$，$\|\boldsymbol{x}_m-\boldsymbol{x}\|\to 0$成立时，称向量序列$\{\boldsymbol{x}_m\}$收敛于$\boldsymbol{x}$，$\boldsymbol{x}$称为该向量序列的极限。

向量序列$\{\boldsymbol{x}_m\}$收敛于$\boldsymbol{x}$，通常用$\lim\limits_{m\to\infty}\boldsymbol{x}_m=\boldsymbol{x}$或$m\to\infty$，$\boldsymbol{x}_m\to\boldsymbol{x}$表示。

定义 8.1.1 给出的只是一般向量空间中收敛的定义，对向量空间F^n（实数域上的向量空间R^n或复数域上的向量空间C^n），可以用向量的分量给出等价的定义。

设$\{\boldsymbol{x}_m\}$是向量空间F^n中的任意向量序列，$\boldsymbol{x}\in F^n$为其极限，它们的分量表示为

$$\boldsymbol{x}_m=\begin{pmatrix}x_1^{(m)}\\x_2^{(m)}\\\vdots\\x_n^{(m)}\end{pmatrix},\qquad \boldsymbol{x}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\tag{8-1}$$

设在F^n中取定范数

$$\|\boldsymbol{x}\|_\infty=\max_i|x_i|\tag{8-2}$$

由式（8-1）和式(8-2)，易知

$$\|\boldsymbol{x}_m-\boldsymbol{x}\|_\infty=\max_i|x_i^{(m)}-x_i|\geqslant|x_i^{(m)}-x_i|\tag{8-3}$$

于是，当$m\to\infty$时

$$\|\boldsymbol{x}_m-\boldsymbol{x}\|\to 0\Rightarrow|x_i^{(m)}-x_i|\to 0\qquad(i=1,\ 2,\ \cdots,\ n)\tag{8-4}$$

反之，若$m\to\infty$，$|x_i^{(m)}-x_i|\to 0$（$i=1$，2，…，n），则

$$\max_i|x_i^{(m)}-x_i|=\|\boldsymbol{x}_m-\boldsymbol{x}\|\to 0$$

成立。于是

$$若m\to\infty,\ \|\boldsymbol{x}_m-\boldsymbol{x}\|\to 0\Leftrightarrow 若m\to\infty,\ |x_i^{(m)}-x_i|\to 0\tag{8-5}$$

由范数的等价性知

① 范数固定的向量空间$V(F)$称为赋范线性空间，记成（$V(F)$，$\|\cdot\|$）等。

$$\text{若} m\to\infty,\ \|x_m - x\|_\infty \to 0 \Leftrightarrow \text{若} m\to\infty,\ \|x_m - x\| \to 0 \qquad (8\text{-}6)$$

因此，F^n 上的向量序列$\{x_m\}$收敛于向量 $x\in F^n$（对于任意范数）与 x_m 的各个分量收敛于 x 的对应分量（在实数或复数数列的意义上）是一样的，于是对 F^n，定义 8.1.1 可叙述为：

定义 8.1.2 在向量空间 F^n中，给定向量序列 $\{x_m\}$，当 x_m 的各个分量收敛时，称$\{x_m\}$收敛，并且称以各分量的极限为分量的向量 $x\in F^n$ 为 $\{x_m\}$ 的极限。

因为 F^n 中向量序列的收敛与各分量的收敛等价，所以可用实数域 R（或复数域 C）上，数列收敛性的判别法则判别其收敛性。例如，由于在 R（或 C）中形成 Cauchy 序列是收敛的充要条件，所以在 F^n 中$\{x_m\}$为 Cauchy 序列[1]也是收敛的充要条件。

须要指出的是，在一般的向量空间$V(F)$中，$\{x_m\}$为 Cauchy 序列是$\{x_m\}$收敛的必要条件而不是充分条件，譬如，定义在$t\in(0,1)$上的实系数多项式的全体$P(t)$，构成 R 上的一个向量空间。对于 $P(t)$ 上的向量序列和范数，考虑

$$x_m(t)=\sum_{k=1}^{m}\frac{t^k}{k!},\quad \|x_m\|=\max_{t\in[0,1]}|x_m(t)|$$

因为

$$\|x_n-x_m\|\leqslant\frac{1}{(m+1)!}+\cdots+\frac{1}{n!}$$

所以$\{x_m\}$是 Cauchy 序列，但是由于当 $m\to\infty$ 时 $x_m(t)\to e^t$，故在 $P(t)$ 中没有极限。

§8.2 矩阵序列与极限

这一节我们讨论矩阵序列的收敛性。

设给出 $F^{n\times n}$ 中的矩阵序列

$$A_1,\ A_2,\ \cdots A_m,\ \cdots$$

用 $\{A_m\}$ 来表示，因为矩阵由有限个元素组成，其收敛性可以和 F^n 中的向量一样地考虑，所以，我们也按照定义 8.1.1 或 8.1.2 规定矩阵序列的收敛性。

定义 8.2.1 在 $F^{n\times n}$ 的矩阵序列 $\{A_m\}$ 中，令$A_m=(a_{ij}^{(m)})_{n\times n}$，如果 n^2 个数列$\{a_{ij}^{(m)}\}$，$(i,\ j=1,\ 2,\ \cdots,\ n)$都收敛时，称 $\{A_m\}$ 收敛。以数列 $\{a_{ij}^{(m)}\}$ 的极限为元素的矩阵 $A\in F^{n\times n}$，称为 $\{A_m\}$ 的极限，记作

$$\lim_{m\to\infty} A_m=A,\ \text{或} m\to\infty,\ A_m\to A$$

下面我们来证明，定义8.2.1 和条件

$$m\to\infty,\ \|A_m-A\|\to 0 \qquad (8\text{-}7)$$

（其中$\|\cdot\|$为任一矩阵范数）是等价的。

取矩阵范数

---

[1] 在赋范(线性)空间中，当 $n,m\to\infty$ 时，$\|x_n-x_m\|\to 0$ 的向量序列$\{x_m\}$称为 Cauchy序列

$$\|A\| = \sum_{i,j=1}^{n} |a_{ij}|$$

设 $\lim_{m \to \infty} A_m = A$。于是对每个 i，有

$$\lim_{m \to \infty} a_{ij}^{(m)} = a_{ij} \qquad (i,\ j=1,\ 2\cdots,\ n)$$

即

$$\lim_{m \to \infty} |a_{ij}^{(m)} - a_{ij}| = 0 \quad (i,\ j=1,2,\cdots,\ n)$$

故有

$$\lim_{m \to \infty} \|A_m - A\| = \lim_{m \to \infty} \sum_{i,j=1}^{n} |a_{ij}^{(m)} - a_{ij}| = 0$$

$$(i,\ j=1,\ 2,\ \cdots,\ n)$$

反之，设 $\lim_{m \to \infty} \|A_m - A\| = 0$，即

$$\lim_{m \to \infty} \sum_{i,j=1}^{n} |a_{ij}^{(m)} - a_{ij}| = 0$$

于是对 i，$j=1$，2，$\cdots n$，有

$$\lim_{m \to \infty} |a_{ij}^{(m)} - a_{ij}| = 0$$

即

$$\lim_{m \to \infty} A_m = A$$

由矩阵范数的等价性

$$c_1 \|A_m - A\| \leqslant \|A_m - A\|_a \leqslant c_2 \|A_m - A\|$$

$$c_1' \|A_m - A\|_a \leqslant \|A_m - A\| \leqslant c_2' \|A_m - A\|_a$$

其中 $\|\cdot\|_a$ 为任一矩阵范数，知

$$\lim_{m \to \infty} \|A_m - A\| = 0 \Leftrightarrow \lim_{m \to \infty} \|A_m - A\|_a = 0$$

利用数列收敛的概念和性质容易验证：

（1） 一个收敛矩阵序列的极限是唯一的。

（2） 设 $\lim_{m \to \infty} A_m = A$，$\lim_{m \to \infty} B_m = B$，则

$$\lim_{m \to \infty} (aA_m + bB_m) = aA + bB \qquad a,\ b \in F \tag{8-8}$$

（3） 设 $\lim_{m \to \infty} A_m = A$，$\lim_{m \to \infty} B_m = B$，$A_m$，$B_m \in F^{n \times n}$，则

$$\lim_{m \to \infty} A_m B_m = AB \tag{8-9}$$

（4） 设 $\lim_{m \to \infty} A_m = A$，$A_m$、$P$、$Q \in F^{n \times n}$，则

$$\lim_{m\to\infty} PA_mQ = PAQ \tag{8-10}$$

（5） 设 $\lim_{m\to\infty} A_m = A$，且 $A_m(m=1, 2, \cdots)$、A均可逆，则 $\{A_m^{-1}\}$ 也收敛，且

$$\lim_{m\to\infty} A_m^{-1} = A^{-1}$$

〔证明〕 设 A_m^* 为 A_m 的伴随矩阵，则

$$A_m^{-1} = \frac{A_m^*}{|A_m|} \tag{8-11}$$

其中

$$A_m^* = \begin{pmatrix} A_{11}^{(m)} & A_{21}^{(m)} & \cdots & A_{n1}^{(m)} \\ A_{12}^{(m)} & A_{22}^{(m)} & \cdots & A_{n2}^{(m)} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n}^{(m)} & A_{2n}^{(m)} & \cdots & A_{nn}^{(m)} \end{pmatrix}$$

因为 $A_{ij}^{(m)}$ 是 $|A_m|$ 中元素 $a_{ij}^{(m)}$ 的代数余子式，所以 $A_{ij}^{(m)}$ 是 A_m 的元素的（$n-1$）次多项式。由多项式函数的连续性知

$$\lim_{m\to\infty} A_{ij}^{(m)} = A_{ij}$$

其中 A_{ij} 为$|A|$中元素 a_{ij} 的代数余子式。故有

$$\lim_{m\to\infty} A_m^* = A^*$$

又 $|A_m|$ 是 A_m 的元素的 n 次多项式，所以

$$\lim_{m\to\infty} |A_m| = |A| \neq 0$$

于是由式（8-11）有

$$\lim_{m\to\infty} A_m^{-1} = \lim_{m\to\infty} \frac{A_m^*}{|A_m|} = \frac{A^*}{|A|} = A^{-1}$$]

设 $\{A_m\}$ 是由矩阵 $A \in F^{n\times n}$ 的幂组成的矩阵序列：

$$A, A^2, \cdots, A^m, \cdots \tag{8-12}$$

下面我们来证明关于矩阵序列（8-12）的两个定理。

定理 8.2.1 若对矩阵 A 的某一种范数有 $\|A\|<1$，则 $\lim_{m\to\infty} A^m = 0$。

〔证明〕 因为

$$\|A^m\| \leqslant \|A\|^m$$

所以当 $\|A\|<1$ 时，有

$$\lim_{m\to\infty} \|A^m\| = 0$$

即

$$\lim_{m\to\infty}\|A^m-0\|=0$$

由式（8-7）知

$$\lim_{m\to\infty}A^m=0$$

∎

定理 8.2.2 设有矩阵序列 $\{A_m\}$：A，A^2，…，A^m，…，则 $\lim_{m\to\infty}A^m=0$ 的充要条件是 A 的所有特征值的模均小于 1。

〔证明〕 设

$$A\sim J=\mathrm{diag}(J_1(\lambda_1),\ J_2(\lambda_2),\ \cdots,\ J_r(\lambda_r))$$

其中

$$J_i(\lambda_i)=\begin{pmatrix}\lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}_{d_i\times d_i}\quad(i=1,\ 2,\ \cdots,\ r)$$

即存在非奇异矩阵 P，使

$$P^{-1}AP=J,\ A=PJP^{-1}$$

于是

$$\begin{aligned}A^m&=PJ^mP^{-1}\\&=P\mathrm{diag}(J_1^m(\lambda_1),\ J_2^m(\lambda_2),\ \cdots,\ J_r^m(\lambda_r))P^{-1}\end{aligned}$$

显然，$\lim_{m\to\infty}A^m=0$ 的充要条件是 $\lim_{m\to\infty}J_i^m(\lambda_i)=0$，$i=1,\ 2,\ \cdots,\ r$。

又

$$J_i^m(\lambda_i)=\begin{pmatrix}\lambda_i^m & C_m^1\lambda_i^{m-1} & \cdots & C_m^{d_i-1}\lambda_i^{m-d_i+1}\\ & \lambda_i^m & & \\ & & \ddots & \\ & & & \lambda_i^m\end{pmatrix}_{d_i\times d_i}$$

$i=1,\ 2,\ \cdots,\ r$，当 $r<s$ 时，$C_r^s=0$。

于是，$\lim_{m\to\infty}J_i^m(\lambda_i)=0$ 的充要条件是对 $i=1,\ 2,\ \cdots,\ r$，有 $|\lambda_i|<1$。 ∎

我们指出：如果 $\rho(A)=1$，则 $\lim_{m\to\infty}A^m$ 存在的充要条件是 A 的特征值为 1 的初等因子均为一次的。

定理 8.2.3 对矩阵 A 的任何一种范数 $\|A\|$，均有

$$|\lambda_i| \leqslant \|A\| \qquad (i=1, 2, \cdots, n)$$

其中 λ_1, λ_2, …, λ_n 为矩阵 A 的特征值。

〔证明〕 令 $\|A\|=a$。作矩阵

$$B=\frac{1}{a+d}A$$

其中 d 为任意不为零的正数，对矩阵 B，有

$$\|B\|=\left\|\frac{1}{a+d}A\right\|=\frac{1}{a+d}\|A\|=\frac{a}{a+d}<1$$

于是，由定理 8.2.1，有

$$\lim_{n\to\infty}B^n=0$$

再由定理 8.2.2 知，矩阵 B 的所有特征值的模小于 1。

设 A 的特征值为

$$\lambda_1, \lambda_2, \cdots, \lambda_n$$

那么 B 的特征值为

$$\frac{\lambda_1}{a+d}, \quad \frac{\lambda_2}{a+d}, \cdots, \frac{\lambda_n}{a+d}$$

且

$$\left|\frac{\lambda_i}{a+d}\right|<1 \quad (i=1, 2, \cdots, n)$$

即

$$|\lambda_i|<a+d \qquad (i=1, 2, \cdots, n)$$

由 d 的任意性有

$$|\lambda_i| \leqslant \|A\|$$

定理 8.2.3 说明矩阵 A 的谱半径 $\rho(A)$ 不大于矩阵 A 的任何一种范数。

§8.3 矩阵级数

定义 8.3.1 设有矩阵序列

$$A_1, A_2, \cdots, A_m, \cdots, \tag{8-13}$$

其中 $A_m=(a_{ij}^{(m)})\in F^{n\times n}$，称

$$A_1+A_2+\cdots+A_m+\cdots \tag{8-14}$$

为矩阵级数，A_m 称为矩阵级数的一般项。

级数(8-14)前 m 项的和

$$S_m=A_1+A_2+\cdots+A_m \tag{8-15}$$

称为级数（8-14）的**部分和**，它仍是一个 n 阶方阵。

定义 8.3.2 若 $\lim\limits_{m\to} S_m=S$，则称级数（8-14）**收敛**，S 称为**级数的和**，记作

$$S=A_1+A_2+\cdots+A_m+\cdots=\sum_{i=1}^{\infty}A_m \tag{8-16}$$

显然，矩阵级数（8-14）收敛的充要条件是 n^2 个数项级数 $\sum\limits_{m=1}^{\infty}a_{ij}^{(m)}$（$i, j=1, 2, \cdots, n$）都收敛，且和为

$$S=\begin{pmatrix}\sum\limits_{m=1}a_{11}^{(m)} & \sum\limits_{m=1}^{\infty}a_{12}^{(m)} & \cdots & \sum\limits_{m=1}^{\infty}a_{1n}^{(m)}\\ \sum\limits_{m=1}^{\infty}a_{21}^{(m)} & \sum\limits_{m=1}^{\infty}a_{22}^{(m)} & \cdots & \sum\limits_{m=1}^{\infty}a_{2n}^{(m)}\\ \cdots & \cdots & \cdots & \cdots\\ \sum\limits_{m=1}^{\infty}a_{n1}^{(m)} & \sum\limits_{m=1}^{\infty}a_{n2}^{(m)} & \cdots & \sum\limits_{m=1}^{\infty}a_{nn}^{(m)}\end{pmatrix}$$

由矩阵级数收敛的定义易知

（1） 若 $\sum\limits_{m=1}^{\infty}A_m$ 收敛，则 $\lim\limits_{m\to\infty}A_m=0$。

（2） 若 $\sum\limits_{m=1}^{\infty}A_m=S$，$\sum\limits_{m=1}^{\infty}B_m=S'$，则

$$\sum_{m=1}^{\infty}(A_m\pm B_m)=S\pm S'$$

（3） 若 $\sum\limits_{m=1}^{\infty}A_m=S$，则 $\sum\limits_{m=1}^{\infty}kA_m=kS$，$k\in F$。

定义 8.3.3 设矩阵级数

$$A_1+A_2+\cdots+A_m+\cdots$$

其中 $A_m=(a_{ij}^{(m)})\in F^{n\times n}$，如果 n^2 个数项级数

$$a_{ij}^{(1)}+a_{ij}^{(2)}+\cdots+a_{ij}^{(m)}+\cdots$$

（$i, j=1, 2, \cdots, n$）都绝对收敛，则称矩阵级数 **绝对收敛**。

定理 8.3.1 设 $A_m=(a_{ij}^{(m)})\in F$，则矩阵级数 $\sum\limits_{m=1}^{\infty}A_m$ 绝对收敛的充要条件是级数

$\sum_{m=1}^{\infty}\|A_m\|$收敛，其中$\|A_m\|$为$A_m$的任何一种矩阵范数。

〔证明〕 取

$$\|A_m\| = \sum_{i,j=1}^{n} |a_{ij}^{(m)}|$$

对$i, j=1, 2, \cdots, n$，有

$$\|A_m\| \geqslant |a_{ij}^{(m)}|$$

因此，若$\sum_{m=1}^{\infty}\|A_m\|$收敛，则$n^2$个级数$\sum_{m=1}^{\infty}|a_{ij}^{(m)}|$都收敛。由定义 8.3.3 知，矩阵级数$\sum_{m=1}^{\infty}A_m$绝对收敛。

反之，如果级数$\sum_{m=1}^{\infty}A_m$绝对收敛，则由定义 8 2.3 知，n^2个正项常数项级数

$$\sum_{m=1}^{\infty}|a_{ij}^{(m)}| \qquad (i, j=1, 2, \cdots, n)$$

均收敛。于是每个级数的部分和均有界，设

$$\sum_{m=1}^{N}|a_{ij}^{(m)}| < M_{ij} \qquad (i, j=1, 2, \cdots, n)$$

其中M_{ij}是与N无关的常数。

令

$$M = \max\{M_{ij}, \ i, j=1, 2, \cdots, n\}$$

则有

$$\sum_{m=1}^{N}|a_{ij}^{(m)}| < M \qquad (i, j=1, 2, \cdots, n)$$

于是

$$\sum_{m=1}^{N}\|A_m\| = \sum_{m=1}^{N}\sum_{i,j=1}^{n}|a_{ij}^{(m)}|$$

$$= \sum_{i,j=1}^{n}\sum_{m=1}^{N}|a_{ij}^{(m)}| < n^2 M$$

故$\sum_{m=1}^{\infty}\|A_m\|$收敛。

根据范数等价性定理知，对任河一种矩阵范数，定理均成立。 】

由定理 8.3.1 可以给出与定义 8.3.3 等价的矩阵级数绝对收敛的定义。

定理 8.3.2 设矩阵级数 $\sum\limits_{m=1}^{\infty}A_m$ 绝对收敛。则

（1） 级数 $\sum\limits_{m=1}^{\infty}A_m$ 收敛。

（2） 级数 $\sum\limits_{m=1}^{\infty}A_m$ 在任意改变各项的次序后仍然收敛，且其和不变。

这个定理的证明留给读者。

定理 8.3.3 设 P、Q 为 n 阶非奇异矩阵，若矩阵级数 $\sum\limits_{m=1}^{\infty}A_m$ 收敛（绝对收敛），则矩阵级数 $\sum\limits_{m=1}^{\infty}PA_mQ$ 也收敛（绝对收敛），且

$$\sum_{m=1}^{\infty}PA_mQ=P\left(\sum_{m=1}^{\infty}A_m\right)Q \tag{8-17}$$

〔证明〕 设

$$\sum_{m=1}^{\infty}A_m=S,\qquad \sum_{m=1}^{N}A_m=S_N$$

则

$$\lim_{N\to\infty}S_N=S$$

由式（8-10），有

$$\begin{aligned}\lim_{N\to\infty}PS_NQ&=P(\lim_{N\to\infty}S_N)Q\\&=PSQ=P\left(\sum_{m=1}^{\infty}A_m\right)Q\end{aligned}$$

即当 $\sum\limits_{m=1}^{\infty}A_m$ 收敛时，$\sum\limits_{m=1}^{\infty}PA_mQ$ 也收敛。

若 $\sum\limits_{m=1}^{\infty}A_m$ 绝对收敛，由定理 8.3.1 知，$\sum\limits_{m=1}^{\infty}\|A_m\|$ 也绝对收敛。由矩阵范数的相容性，有

$$\|PA_mQ\|\leqslant\|P\|\|Q\|\|A_m\|=M\|A_m\|$$

其中 $M=\|P\|\cdot\|Q\|$，由此可知，$\sum\limits_{m=1}^{\infty}\|PA_mQ\|$ 收敛，于是 $\sum\limits_{m=1}^{\infty}PA_mQ$ 绝对收敛。 〕

关于两个绝对收敛的矩阵级数，我们给出一个应用起来很方便的结果，即

定理 8.3.4 设 $F^{n\times n}$ 中的两个矩阵级数

$$S_1:\ A_1+A_2+\cdots+A_m+\cdots$$

和

$$S_2:\ B_1+B_2+\cdots+B_m+\cdots$$

都绝对收敛，其和分别为 A、B，则将它们按项相乘后作成的矩阵级数

$$S_3:\ A_1B_1+(A_1B_2+A_2B_1)$$
$$+(A_1B_3+A_2B_2+A_3B_1)+\cdots$$
$$+(A_1B_m+A_2B_{m-1}+\cdots+A_mB_1)+\cdots \tag{8-18}$$

绝对收敛，且具有和AB。

〔证明〕 由定理的假设知，级数$\sum\limits_{m=1}^{\infty}\|A_m\|$，$\sum\limits_{m=1}^{\infty}\|B_m\|$收敛，对于绝对收敛的数项级数，按项相乘作成的级数也绝对收敛。因此，级数

$$\|A_1\|\cdot\|B_1\|+(\|A_1\|\cdot\|B_2\|$$
$$+\|A_2\|\cdot\|B_1\|)+\cdots+(\|A_1\|\cdot\|B_m\|$$
$$+\|A_2\|\cdot\|B_{m-1}\|+\cdots+\|A_m\|\cdot\|B_1\|)+\cdots \tag{8-19}$$

收敛，将式（8-18）和式（8-19）两个级数的各项进行比较知，级数（8-18）绝对收敛。

又设级数 S_1，S_2，S_3 的前 m 项的部分和分别为 S_{1m}，S_{2m}，S_{3m}；级数$\sum\limits_{m=1}^{\infty}\|A_m\|$，$\sum\limits_{m=1}^{\infty}\|B_m\|$的前 m 项的部分和分别为 $\bar{S}_{1m}$，$\bar{S}_{2m}$，级数（8-19）的前 m 项的部分和为 $\bar{S}_{3m}$，则

$$S_{1m}S_{2m}=(A_1B_1+A_1B_2+\cdots+A_1B_m)$$
$$+(A_2B_1+A_2B_2+\cdots+A_2B_m)+\cdots$$
$$+(A_mB_1+A_mB_2+\cdots+A_mA_m)$$
$$S_{3m}=(A_1B_1+A_1B_2+\cdots+A_1B_m)$$
$$+(A_2B_1+A_2B_2+\cdots+A_2B_{m-1})$$
$$+\cdots+A_mB_1$$

因为 $\bar{S}_{1m}\bar{S}_{2m}$，$\bar{S}_{3m}$ 分别是在上面两个级数中，作置换

$$A_i\to\|A_i\|,\ B_i\to\|B_i\|$$

得到的，所以

$$\|S_{1m}S_{2m}-S_{3m}\|\leqslant\bar{S}_{1m}\bar{S}_{2m}-\bar{S}_{3m} \tag{8-20}$$

在式（8-20）中，令 $m\to\infty$，则右边趋向于零，故得

$$\lim_{m\to\infty} S_{1m}S_{2m}=\lim_{m\to\infty}S_{1m}\lim_{m\to\infty}S_{2m}=AB=\lim_{m\to\infty}S_{3m}$$]

例8.3.1 对于$A\in F^{n\times n}$，考虑级数

$$I+A+\frac{A^2}{2!}+\cdots+\frac{A^k}{k!}+\cdots \tag{8-21}$$

因为

$$\|I\|+\|A\|+\frac{\|A\|^2}{2!}+\cdots+\frac{\|A\|^k}{k!}+\cdots=e^{\|A\|}-1+\|I\|$$

所以该级数对于任意$A\in F^{n\times n}$绝对收敛，其和称为A的矩阵指数函数。记作e^A，即

$$e^A=I+A+\frac{A^2}{2!}+\cdots+\frac{A^k}{k!}+\cdots \tag{8-22}$$

最后，我们给出矩阵级数绝对收敛的比较判别法。

定理 8.3.5 设$\sum_{m=1}^{\infty}A_m$，$\sum_{m=1}^{\infty}B_m$是两个矩阵级数，且满足

（1） $\sum_{m=1}^{\infty}B_m$中每一个矩阵$B_m=(b_{ij}^{(m)})$的所有元素$b_{ij}^{(m)}\geqslant 0$。

（2） 矩阵$A_m=(a_{ij}^{(m)})$与$B_m=(b_{ij}^{(m)})$的元素间满足

$$|a_{ij}^{(m)}|\leqslant b_{ij}^{(m)}\quad (i,\ j=1,\ 2,\ \cdots n;\ m=1,\ 2,\ \cdots)$$

（3） 级数$\sum_{m=1}^{\infty}B_m$收敛，则级数$\sum_{m=1}^{\infty}A_m$绝对收敛。

〔证明〕 由条件（2）知

$$\|A_m\|_1=\max_j\left(\sum_{i=1}^{n}|a_{ij}^{(m)}|\right)$$

$$\leqslant\max_j\left(\sum_{i=1}^{n}b_{ij}^{(m)}\right)=\|B\|_1$$

由条件（1）和（3）知，级数$\sum_{m=1}^{\infty}\|B\|_1$收敛，于是级数$\sum_{m=1}^{\infty}\|A\|_1$收敛，从而$\sum_{m=1}^{\infty}A_m$绝对收敛。]

§8.4 矩阵幂级数

定义8.4.1 形如

$$c_0I+c_1A+c_2A^2+\cdots+c_kA^k+\cdots \tag{8-23}$$

的矩阵级数称为矩阵幂级数，其中 $c_i\in F$，$A\in F^{n\times n}$。

因为数项级数

$$\|c_0I\|+\|c_1A\|+\|c_2A^2\|+\cdots+\|c_kA^k\|+\cdots$$

的每一项不大于数项级数

$$|c_0|\|I\|+|c_1|\|A\|+|c_2|\|A\|^2+\cdots$$
$$+|c_k|\|A\|^k+\cdots \tag{8-24}$$

的对应项，所以，若级数（8-24）收敛，则级数（8-23）绝对收敛，于是我们得到：

定理 8.4.1 若正项级数

$$|c_0|\|I\|+|c_1|\|A\|+|c_2|\|A\|^2+\cdots$$
$$+|c_k|\|A\|^k+\cdots$$

收敛，则矩阵幂级数

$$c_0I+c_1A+c_2A^2+\cdots+c_kA^k+\cdots$$

绝对收敛，其中$\|A\|$为矩阵 A的某种范数。

推论 8.4.1 若矩阵 A 的某一种范数$\|A\|$在幂级数

$$c_0+c_1x+c_2x^2+\cdots+c_kx^k+\cdots$$

的收敛圆内，则矩阵幂级数

$$c_0I+c_1A+c_2A^2+\cdots+c_kA^k+\cdots$$

绝对收敛。

例 8.4.1 设

$$A=\begin{pmatrix} 0.2 & 0.5 & 0.1 \\ 0.1 & 0.5 & 0.3 \\ 0.2 & 0.4 & 0.2 \end{pmatrix}$$

则矩阵幂级数

$$I+A+A^2+\cdots+A^k+\cdots \tag{8-25}$$

绝对收敛。

因为幂级数

$$1+x+x^2+\cdots+x^k+\cdots$$

的收敛半径为 1，而

$$\|A\|_\infty=\max_i\left(\sum_j|a_{ij}|\right)=0.9<1$$

故矩阵幂级数（8-25）绝对收敛。

定理8.4.2 设$A\in F^{n\times n}$，如果A的所有特性值都在幂级数

$$\sum_{k=0}^{\infty}c_k x^k \tag{8-26}$$

的收敛圆内，那么矩阵幂级数

$$\sum_{k=0}^{\infty}c_k A^k \tag{8-27}$$

绝对收敛。如果A的特征值中有一个在幂级数（8-26）的收敛圆外，则矩阵幂级数（8-27）发散。

〔证明〕 设J为A的Jordan标准形，则

$$A=PJP^{-1}=P(J_1\dotplus J_2\dotplus\cdots\dotplus J_r)P^{-1}$$

其中

$$J_i=\begin{pmatrix}\lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}_{d_i\times d_i} \tag{8-28}$$

因此

$$A^k=PJ^kP^{-1}=P(J_1^k\dotplus J_2^k\dotplus\cdots\dotplus J_r^k)P^{-1}$$

于是

$$\sum_{k=0}^{\infty}c_kA^k=P\sum_{k=0}^{\infty}(J_1^k\dotplus J_2^k\dotplus\cdots\dotplus J_r^k)P^{-1}$$

而

$$\sum c_kJ_i^k=\begin{pmatrix}\sum c_k\lambda_i^k & \sum c_kC_k^1\lambda_i^{k-1} & \sum c_kC_k^2\lambda_i^{k-2} & \cdots & \sum c_kC_k^{d_i-1}\lambda_i^{k-d_i+1}\\ & \sum c_k\lambda_i^k & \sum c_kC_k^1\lambda_i^{k-1} & \cdots & \sum c_kC_k^{d_i-2}\lambda_i^{k-d_i+2}\\ & & \ddots & & \\ & & & \ddots & \\ & & & & \ddots \\ & & & & \sum c_k\lambda_i^k\end{pmatrix} \tag{8-29}$$

其中

$$\begin{cases} C_k^i = \dfrac{k(k-1)\cdots(k-i+1)}{i!} & \text{当 } i \geqslant 1 \text{ 时} \\ C_k^0 = 0 \end{cases}$$

令 R 为幂级数 $\sum\limits_{k=0}^{\infty} c_k x^k$ 的收敛半径，当 $|\lambda_i| > R$ 时，$\sum\limits_{k=0}^{\infty} c_k J_i^k$ 发散；当 $|\lambda_i| < R$ 时，级数

$$\sum_{k=0}^{\infty} c_k \lambda_i^k, \quad \sum_{k=0}^{\infty} c_k k \lambda_i^{k-1}, \quad \cdots, \quad \sum_{k=0}^{\infty} c_k k(k-1)\cdots(k-d_i+1)\lambda_i^{k-d_i+1}$$

均绝对收敛。因此，级数

$$\sum_{k=0}^{\infty} c_k \lambda_i^k, \quad \sum_{k=0}^{\infty} c_k C_k^1 \lambda_i^{k-1}, \quad \cdots, \quad \sum_{k=0}^{\infty} c_k C_k^{d_i-1} \lambda_i^{k-d_i+1}$$

也都绝对收敛。于是，式（8-29）右端的所有级数都绝对收敛。故当所有的 $|\lambda_i| < R$ 时，$\sum\limits_{k=0}^{\infty} c_k J_i^k$ 绝对收敛，从而级数 $\sum\limits_{k=0}^{\infty} c_k A^k$ 绝对收敛。]

定理 8.4.3 矩阵幂级数

$$I + A + A^2 + \cdots + A^k + \cdots \tag{8-30}$$

绝对收敛的充要条件是 A 的所有特征值的模小于1，且其和为 $(I-A)^{-1}$。

〔证明〕 因为幂级数

$$1 + x + x^2 + \cdots + x^k + \cdots$$

的收敛半径为1，所以当 A 的所有特征值的模均小于1时，由定理 8.4.2 知，级数（8-30）绝对收敛。

反之，若级数（8-30）绝对收敛，则

$$\lim_{k\to\infty} A^k = 0$$

由定理 8.2.2 知，A 的所有特征值的模小于1。

下面来计算级数（8-30）的和。

因为级数（8-30）绝对收敛，而 $I-A$ 本身作为矩阵级数也绝对收敛。由定理 8.3.4 知，

$$(I-A)(I+A+A^2+\cdots+A^k+\cdots) = I$$

由逆矩阵的唯一性，有

$$(I-A)^{-1} = I + A + A^2 + \cdots + A^k + \cdots$$]

推论 8.4.1 设 $\|A\| < 1$，则 $I-A$ 与 $I+A$ 均可逆。

〔证明〕 当$\|A\|<1$时，由例 8.2.1 知，A 的所有特征值的模都小于 1，因而由定理 8.4.3 可知，$(I-A)^{-1}$ 存在。又因$\|-A\|=\|A\|<1$，故$(I+A)^{-1}=(I-(-A))^{-1}$亦存在。]

定理 8.4.4 设$\|A\|<1$，则对任何非负整数 k，有

$$\|(I-A)^{-1}-(I+A+A^2+\cdots+A^k)\|\leqslant\frac{\|A\|^{k+1}}{1-\|A\|} \tag{8-31}$$

〔证明〕 因为$\|A\|<1$，由推论 8.4.1 知，$(I-A)^{-1}$ 存在，又因为对任何非负整数 k，有

$$\begin{aligned}I-[&(I+A+A^2+\cdots+A^k)\\&+(A^{k+1}+A^{k+2}+\cdots+A^{k+l})](I-A)\\&\quad=A^{k+l+1}\end{aligned}$$

故

$$\begin{aligned}&(I-A)^{-1}(I+A+A^2+\cdots+A^k)\\&\qquad=A^{k+1}+A^{k+2}+\cdots+A^{k+l}+A^{k+l+1}(I-A)^{-1}\end{aligned}$$

记

$$B_l=A^{k+1}+A^{k+2}+\cdots+A^{k+l}$$

因为

$$\lim_{l\to\infty}A^{k+l+1}(I-A)^{-1}=0$$

所以，有

$$\lim_{l\to\infty}B_l=(I-A)^{-1}-(I+A+A^2+\cdots+A^k)$$

从而，有

$$\|(I-A)^{-1}-(I+A+A^2+\cdots+A^k)\|=\lim_{l\to\infty}\|B_l\|$$

但

$$\begin{aligned}\|B_l\|&\leqslant\|A^{k+1}\|+\|A^{k+2}\|+\cdots+\|A^{k+l}\|\\&\leqslant\|A\|^{k+1}+\|A\|^{k+2}+\cdots+\|A\|^{k+l}\\&=\frac{\|A\|^{k+1}(1-\|A\|^l)}{1-\|A\|}\end{aligned}$$

故对一切正整数 l，有

$$\|B_l\|\leqslant\frac{\|A\|^{k+1}}{1-\|A\|}$$

于是

$$\|(I-A)^{-1}-(I+A+A^2+\cdots+A^k)\|\leqslant\frac{\|A\|^{k+1}}{1-\|A\|}$$ ▌

我们知道，幂级数

$$1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots+\frac{x^k}{k!}+\cdots$$

$$1-\frac{x^2}{2!}+\frac{x^4}{4!}-\cdots+(-1)^k\frac{x^{2k}}{(2k)!}+\cdots$$

$$x-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots+(-1)^k\frac{x^{2k+1}}{(2k+1)!}+\cdots$$

的收敛半径均为∞，由上面的讨论知，矩阵幂级数

$$I+A+\frac{A^2}{2!}+\cdots+\frac{A^k}{k!}+\cdots$$

$$I-\frac{A^2}{2!}-\frac{A^4}{4!}-\cdots+(-1)^k\frac{A^{2k}}{(2k)!}+\cdots$$

$$A-\frac{A^3}{3!}+\frac{A^5}{5!}-\cdots+(-1)^k\frac{A^{2k+1}}{(2k+1)!}+\cdots$$

均绝对收敛。

作为向量、矩阵的极限和范数的一个应用我们来讨论求解线性代数方程组迭代法的收敛性问题。

对于$n\times n$线性代数方程组

$$A\mathbf{Z}=\boldsymbol{b}$$

如果解向量是$\mathbf{Z}^*$，迭代法的思想是构造出一个向量序列$\{\mathbf{Z}_k\}$，使得

$$\lim_{k\to\infty}\mathbf{Z}_k=\mathbf{Z}^*$$

如何具体构造$\{\mathbf{Z}_k\}$是线性代数计算方法所讨论的内容。这里我们只指出，构造$\{\mathbf{Z}_k\}$的方法都可以归结为如下形式

$$\mathbf{Z}_{k+1}=B\mathbf{Z}_k+\boldsymbol{d}\qquad(k=0,\ 1,\ 2,\ \cdots)\tag{8-32}$$

其中B为n阶方阵，$\boldsymbol{d}$是n维向量，并且使得$(I-B)^{-1}$存在，而

$$\mathbf{Z}^*=B\mathbf{Z}^*+\boldsymbol{d}$$

由式（8-32）构造$\{\mathbf{Z}_k\}$，只要预先给定一个向量$\mathbf{Z}_0$称为初始向量，依次可得

$$\mathbf{Z}_1=B\mathbf{Z}_0+\boldsymbol{d}$$

$$\mathbf{Z}_2=B\mathbf{Z}_1+\boldsymbol{d}$$

$$\cdots\cdots\cdots\cdots$$

这样，每一个$\mathbf{Z}_k$都可以构造出来。

下面我们来讨论，矩阵 B 满足什么样的条件，才能保证序列 $\{\boldsymbol{Z}_k\}$ 总是收敛到 $\boldsymbol{Z}^*$。

定理 8.4.5 对任意的初始向量 $\boldsymbol{Z}_0$ 和任意的自由项 $\boldsymbol{d}$，由式（8-32）确定的向量序列 $\{\boldsymbol{Z}_k\}$ 收敛于 $\boldsymbol{Z}^*$ 的充要条件是矩阵 B 的所有特征值的模都小于 1。

〔证明〕 充分性 设 B 的所有特征值的模均小于 1。依次将 $\boldsymbol{Z}_1$ 代入 $\boldsymbol{Z}_2$，$\boldsymbol{Z}_2$ 代入 $\boldsymbol{Z}_3$，…，可得

$$\begin{cases} \boldsymbol{Z}_1 = B\boldsymbol{Z}_0 + \boldsymbol{d} \\ \boldsymbol{Z}_2 = B\boldsymbol{Z}_1 + \boldsymbol{d} = B^2\boldsymbol{Z}_0 + B\boldsymbol{d} + \boldsymbol{d} \\ \boldsymbol{Z}_3 = B\boldsymbol{Z}_2 + \boldsymbol{d} = B^3\boldsymbol{Z}_0 + B^2\boldsymbol{d} + B\boldsymbol{d} + \boldsymbol{d} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \boldsymbol{Z}_k = B^k\boldsymbol{Z}_0 + B^{k-1}\boldsymbol{d} + B^{k-2}\boldsymbol{d} + \cdots + B\boldsymbol{d} + \boldsymbol{d} \\ \quad = B^k\boldsymbol{Z}_0 + (I + B + B^2 + \cdots + B^{k-1})\boldsymbol{d} \end{cases} \tag{8-33}$$

由假设和定理 8.2.2，有

$$\lim_{k\to\infty} B^k = 0$$

再由定理 8.4.3，有

$$I + B + B^2 + \cdots + B^k + \cdots = (I - B)^{-1}$$

于是由式（8-33），得

$$\lim_{k\to\infty} \boldsymbol{Z}_k = (I - B)^{-1}\boldsymbol{d} = \boldsymbol{Z}^*$$

即

$$\boldsymbol{Z}^* = B\boldsymbol{Z}^* + \boldsymbol{d}$$

必要性 因为 $\boldsymbol{d}$ 是任意的，故可取 $\boldsymbol{d}=0$，这时有 $\boldsymbol{Z}^*=0$。由式（8-33），有

$$\boldsymbol{Z}_k = B^k\boldsymbol{Z}_0 \tag{8-34}$$

由式（8-34），有

$$\lim_{k\to\infty} B^k = 0$$

假如不然，B^k 中至少有一个元素 $b_{ij}^{(k)}$ 不收敛于零。取向量

$$\boldsymbol{Z}_0 = (0,\ 0,\ \cdots,\ 0,\ 1,\ 0,\ \cdots,\ 0)$$

于是向量

$$\boldsymbol{Z}_k = B^k\boldsymbol{Z}_0 = (b_{1j}^{(k)},\ b_{2j}^{(k)},\ \cdots,\ b_{ij}^{(k)},\ \cdots,\ b_{nj}^{(k)})$$

不收敛于零。此与必要性的假设矛盾。于是由定理 8.2.2 知，B 的所有特征值的模都小于 1。 ▮

推论 8.4.2 如果对某种与向量相容的范数，$\|B\|<1$，那么序列 $\{\boldsymbol{Z}_k\}$ 收敛到 $\boldsymbol{Z}^*$，且

有误差估计式

$$\|\boldsymbol{Z}^*-\boldsymbol{Z}_k\|\leqslant\frac{\|B\|^k}{1-\|B\|}(\|\boldsymbol{d}\|+\|\boldsymbol{Z}_0\|) \tag{8-35}$$

〔证明〕 推论前一部分的结论由定理 8.4.5 可以直接得出，下面证明式（8-35）。

因为

$$\boldsymbol{Z}^*=(I-B)^{-1}\boldsymbol{d}$$

而

$$\boldsymbol{Z}^*-\boldsymbol{Z}_k=(I-B)^{-1}\boldsymbol{d}-(I+B+B^2+\cdots+B^{k-1})\boldsymbol{d}-B^k\boldsymbol{Z}_0$$

$$=[(I-B)^{-1}-(I+B+B^2+\cdots+B^{k-1})]\boldsymbol{d}-B^k\boldsymbol{Z}_0$$

于是由定理 8.4.4，有

$$\|\boldsymbol{Z}^*-\boldsymbol{Z}_k\|\leqslant\frac{\|B\|^k}{1-\|B\|}\|\boldsymbol{d}\|+\|B\|^k\|\boldsymbol{Z}_0\|$$

$$\leqslant\frac{\|B\|^k}{1-\|B\|}(\|\boldsymbol{d}\|+\|\boldsymbol{Z}_0\|)$$ ▌

例8.4.2 对于方程组

$$\begin{cases}5x_1+2x_2=8\\3x_1-20x_2=26\end{cases}$$

可以改写成

$$\begin{cases}x_1=-0.4x_2+1.6\\x_2=0.15x_1-1.3\end{cases}$$

或

$$\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}0&-0.4\\0.15&0\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}1.6\\-1.3\end{pmatrix}$$

记

$$\boldsymbol{Z}_k=\begin{pmatrix}x_1^{(k)}\\x_2^{(k)}\end{pmatrix},\quad B=\begin{pmatrix}0&-0.4\\0.15&0\end{pmatrix},\quad \boldsymbol{d}=\begin{pmatrix}1.6\\-1.3\end{pmatrix}$$

取 $\boldsymbol{Z}_0=(0,\ 0)^T$ 由

$$\boldsymbol{Z}_k=B\boldsymbol{Z}_{k-1}+\boldsymbol{d}$$

得到 $\boldsymbol{Z}_k$:

k	$x_1^{(k)}$	$x_2^{(k)}$
0	0	0
1	1.6	−1.3
2	2.12	−1.6
3	2.024	−0.982
4	1.9928	−0.9964
5	1.99856	−1.00108
6	2.000432	−1.000216

由向量范数$\|\cdot\|_\infty$，有

$$\|\boldsymbol{d}\|_\infty=1.6,\ \|\boldsymbol{Z}_0\|_\infty=0,\ \|B\|_\infty=0.4$$

于是有估计式

$$\|\boldsymbol{Z}^*-\boldsymbol{Z}_6\|\leqslant\frac{(0.4)^6\times1.6}{1-0.4}\leqslant0.011$$

习 题

8-1 若$\lim\limits_{m\to\infty}A_m=A$，$\lim\limits_{m\to\infty}B_m=B$，证明

$$\lim_{m\to\infty}(aA_m+bB_m)=aA+bB$$

其中a、b为常数。

8-2 若 $\lim\limits_{m\to\infty}A_m=A$，$\lim\limits_{m\to\infty}B_m=B$， P，$Q\in F^{n\times n}$，证明

$$\lim_{m\to}A_mB_m=AB$$

$$\lim_{m\to\infty}PA_mQ=PAQ$$

8-3 证明：对任意$A\in F^{n\times n}$

$$\cos A=I-\frac{A^2}{2!}+\frac{A^4}{4!}-\cdots+(-1)^k\frac{A^{2k}}{(2k)!}+\cdots$$

$$\sin A=A-\frac{A^3}{3!}+\frac{A^5}{5!}-\cdots+(-1)^k\frac{A^{2k+1}}{(2k+1)!}+\cdots$$

绝对收敛。

8-4 证明定理 8.3.2。

8-5 证明推论 8.4.2。

8-6 对$A\in F^{n\times n}$，设 $\sin A$ 和$\cos A$ 由题 8-3 定义，试证明下面等式

（1） $\cos^2 A+\sin^2 A=I$

（2） 当 $AB=BA$ 时，

$$\sin(A+B)=\sin A\cos B+\sin B\cos A$$

$$\cos(A+B)=\cos A\cos B-\sin A\sin B$$

$$e^{A+iB}=e^{A}(\cos B+i\sin B)$$

第九章 矩阵函数

矩阵函数是矩阵分析中的一个重要内容，它在常微分方程组及系统理论中均有不少的应用。

本章讨论的内容有：矩阵多项式，矩阵函数的定义及性质，矩阵函数的五种表示法。

§9.1 矩阵多项式

定义 9.1.1 设 $A\in C^{n\times n}$，$p(\lambda)=a_m\lambda^m+a_{m-1}\lambda^{m-1}+\cdots+a_1\lambda+a_0$，$(a_i\in C)$ 是一个纯量多项式，用 A^i 代替 λ^i，用 I 代替 $\lambda^0=1$，得到的表达式

$$a_mA^m+a_{m-1}A^{m-1}+\cdots+a_1A+a_0I$$

称为矩阵多项式，记作 $p(A)$，即

$$p(A)=a_mA^m+a_{m-1}A^{m-1}+\cdots+a_1A+a_0I \tag{9-1}$$

若 $a_m\neq 0$，则称 $p(A)$ 为 m 次矩阵多项式。

两个矩阵多项式 $p_1(A)$、$p_2(A)$，它们的和与积分别为

$$p_1(A)+p_2(A)=(p_1+p_2)(A) \tag{9-2}$$

$$p_1(A)\cdot p_2(A)=(p_1\cdot p_2)(A) \tag{9-3}$$

其中 $(p_1+p_2)(A)$，$(p_1p_2)(A)$ 分别表示先按纯量多项式 $p_1(\lambda)$，$p_2(\lambda)$ 求它们的和与积，然后在所得到的多项式中，将 λ^i 换成 A^i，$\lambda^0=1$ 换成 I。

下面来讨论矩阵多项式的性质。

因为纯量多项式的乘法是可换的，所以有

定理 9.1.1 $p_1(A)p_2(A)=p_2(A)p_1(A)$

定理 9.1.2 设 $p(\lambda)=a_m\lambda^m+a_{m-1}\lambda^{m-1}+\cdots+a_1\lambda+a_0$，$A$ 为分块对角矩阵

$$A=\begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_t \end{bmatrix}$$

则

$$p(A)=\begin{bmatrix} p(A_1) & & & \\ & p(A_2) & & \\ & & \ddots & \\ & & & p(A_t) \end{bmatrix}$$

〔证明〕 因为

$$A^i=\begin{bmatrix} A_1^i & & & \\ & A_2^i & & \\ & & \ddots & \\ & & & A_t^i \end{bmatrix} \quad (i=0,1,2,\cdots,m)$$

所以有

$$p(A)=a_mA^m+a_{m-1}A^{m-1}+\cdots+a_1A+a_0I$$

$$=a_m\begin{bmatrix} A_1^m & & & \\ & A_2^m & & \\ & & \ddots & \\ & & & A_t^m \end{bmatrix}+a_{m-1}\begin{bmatrix} A_1^{m-1} & & & \\ & A_2^{m-1} & & \\ & & \ddots & \\ & & & A_t^{m-1} \end{bmatrix}$$

$$+a_1\begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_t \end{bmatrix}+a_0\begin{bmatrix} I_1 & & & \\ & I_2 & & \\ & & \ddots & \\ & & & I_t \end{bmatrix}$$

$$=\begin{bmatrix} \sum_{k=0}^{m}a_kA_1^k & & & \\ & \sum_{k=0}^{m}a_kA_2^k & & \\ & & \ddots & \\ & & & \sum_{k=0}^{m}a_kA_t^k \end{bmatrix}$$

$$=\begin{bmatrix} p(A_1) & & & \\ & p(A_2) & & \\ & & \ddots & \\ & & & p(A_t) \end{bmatrix}$$

▮

定理 9.1.3 设 P 是非奇异矩阵，则

$$p(P^{-1}AP)=P^{-1}p(A)P$$

〔证明〕 由 $(P^{-1}AP)^i=P^{-1}A^iP$，有

$$p(P^{-1}AP)=a_mP^{-1}A^mP+a_{m-1}P^{-1}A^{m-1}P+\cdots+a_1P^{-1}AP+a_0P^{-1}IP$$

$$=P^{-1}(a_mA^m+a_{m-1}A^{m-1}+\cdots+a_1A+a_0I)P$$

$$=P^{-1}p(A)P$$

▮

由定理 9.1.2 和定理 9.1.3 可以得出计算矩阵多项式的一个方法。

设矩阵 A 的 Jordan 标准形为 J，故

$$P^{-1}AP=J$$

其中

$$J=\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{pmatrix}$$

$$J_i=\begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{d_i \times d_i}$$

于是，由定理 9.1.2 和定理 9.1.3，有

$$\begin{aligned} p(A) &= Pp(J)P^{-1} \\ &= P\begin{pmatrix} p(J_1) & & & \\ & p(J_2) & & \\ & & \ddots & \\ & & & p(J_r) \end{pmatrix}P^{-1} \end{aligned} \tag{9-4}$$

而

$$J_i^k=\begin{pmatrix} \lambda_i^k & C_k^1\lambda_i^{k-1} & C_k^2\lambda_i^{k-2} & \cdots & C_k^{d_i-1}\lambda_i^{k-d_i+1} \\ & \lambda_i^k & C_k^1\lambda_i^{k-1} & \cdots & C_k^{d_i-2}\lambda_i^{k-d_i+2} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & C_k^1\lambda_i^{k-1} \\ & & & & \lambda_i^k \end{pmatrix}_{d_i \times d_i}$$

不难验证

$$p(J_i)=\begin{pmatrix} p(\lambda_i) & p'(\lambda_i) & \frac{1}{2!}p''(\lambda_i) & \cdots & \frac{p^{(d_i-1)}(\lambda_i)}{(d_i-1)!} \\ & p(\lambda_i) & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & p'(\lambda_i) \\ & & & & p(\lambda_i) \end{pmatrix} \tag{9-5}$$

其中$p'(\lambda_i)$，$p''(\lambda_i)$，…，$p^{(d_i-1)}(\lambda_i)$分别为$p(\lambda)$在$\lambda=\lambda_i$的各阶导数值。把$p(J_i)$代入式（9-4），便可求出$p(A)$。

例 9.1.1 设

$$A=\begin{pmatrix}2 & 0 & 0\\ 1 & 1 & 1\\ 1 & -1 & 3\end{pmatrix}$$

$$p(\lambda)=\lambda^4-2\lambda^3+\lambda-1$$

计算$p(A)$。

〔解〕 矩阵A的 Jordan 标准形为

$$J=\begin{pmatrix}2 & 1 & 0\\ 0 & 2 & 0\\ 0 & 0 & 2\end{pmatrix}$$

$$T=\begin{pmatrix}0 & 1 & 1\\ 1 & 0 & 0\\ 1 & 0 & -1\end{pmatrix},\quad T^{-1}=\begin{pmatrix}0 & 1 & 0\\ 1 & -1 & 1\\ 0 & 1 & -1\end{pmatrix}$$

$$J_1=\begin{pmatrix}2 & 1\\ 0 & 2\end{pmatrix},\qquad J_2=(2)$$

因而 $$p(J_1)=\begin{pmatrix}p(2) & p'(2)\\ 0 & p(2)\end{pmatrix}=\begin{pmatrix}1 & 9\\ 0 & 1\end{pmatrix}$$

$$p(J_2)=(p(2))=(1)$$

于是 $$p(A)=Pp(J)P^{-1}$$

$$=\begin{pmatrix}0 & 1 & 0\\ 0 & -1 & 1\\ 0 & 1 & -1\end{pmatrix}\begin{pmatrix}[illegible] & [illegible] & [illegible]\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}[illegible] & [illegible] & [illegible]\\ 1 & 0 & 0\\ 1 & 0 & -1\end{pmatrix}$$

$$=\begin{pmatrix}1 & 0 & 0\\ 9 & -8 & 9\\ 8 & -9 & 10\end{pmatrix}$$

在矩阵多项式的讨论中，矩阵的化零多项式和最小多项式占有重要的地位。

定义 9.1.2 设$A\in C^{n\times n}$，满足

$$p(A)=0$$

的多项式 $p(\lambda)$ 称为 A 的**化零多项式**。

在 A 的化零多项式中，次数最低且首项系数为 1 的叫做 A 的**最小多项式**，记作 $\psi_m(\lambda)$。

定理 9.1.4 （Cayley-Hamilton 定理） 设 A 是 n 阶矩阵，且

$$D(\lambda)=\det(\lambda I-A)$$

$$=\lambda^n+c_{n-1}\lambda^{n-1}+\cdots+c_1\lambda+c_0$$

是 A 的特征多项式，则

$$D(A)=0$$

即

$$A^n+c_1A^{n-1}+\cdots+c_1A+c_0I=0 \tag{9-6}$$

〔证明〕 设 B 为矩阵 $\lambda I-A$ 的伴随矩阵，则

$$(\lambda I-A)B=D(\lambda)I \tag{9-7}$$

又 B 的元素是 $\lambda I-A$ 的 $n-1$ 阶余子式，所以它是 λ 的多项式，且次数最多是 $n-1$。因此，方阵 B 的每个元素都是 λ 的多项式，且次数最多是 $n-1$。由矩阵加法的定义，有

$$B=\lambda^{n-1}B_{n-1}+\lambda^{n-2}B_{n-2}+\cdots+\lambda B_1+B_0$$

其中 B_{n-1}、B_{n-2}、…、B_1、B_0 均为方阵，又 $B_{n-1}\neq 0$ 。将上式代入式（9-7），有

$$(\lambda I-A)(\lambda^{n-1}B_{n-1}+\lambda^{n-2}B_{n-2}+\cdots+\lambda B_1+B_0)$$

$$=(\lambda^n+c_{n-1}\lambda^{n-1}+\cdots+c_1\lambda+c_0)I$$

把上式改写成

$$\lambda^nB_{n-1}+\lambda^{n-1}(B_{n-2}-AB_{n-1})+\cdots+\lambda(B_0-AB_1)-AB_0$$

$$=(\lambda^n+c_{n-1}\lambda^{n-1}+\cdots+c_1\lambda+c_0)I$$

比较上式两端相同项的系数，有

$$\begin{cases} B_{n-1}=I \\ B_{n-2}-AB_{n-1}=c_{n-1}I \\ B_{n-3}-AB_{n-2}=c_{n-2}I \\ \cdots\cdots\cdots\cdots \\ B_0-AB_1=c_1I \\ \quad -AB_0=c_0I \end{cases}$$

于是

$$\begin{cases} B_{n-1}=I \\ B_{n-2}=c_{n-1}I+AB_{n-1}=c_{n-1}I+A \\ B_{n-3}=c_{n-2}I+AB_{n-2}=c_{n-2}I+c_{n-1}A+A^2 \\ \cdots\cdots\cdots\cdots \\ B_0=A^{n-1}+c_{n-1}A^{n-2}+\cdots+c_2A+c_1I \end{cases}$$

再由 $-AB_0=c_0I$，有

$$A(A^{n-1}+c_{n-1}A^{n-2}+\cdots+c_2A+c_1I)+c_0I=0$$

即
$$A^n+c_{n-1}A^{n-1}+\cdots+c_1A+c_0I=0$$
▮

定理 9.1.5 A 的任意化零多项式都可以用 A 的最小多项式整除。

〔证明〕 我们知道，对于任意次数大于最小多项式 $\psi_m(\lambda)$ 次数的多项式 $p(\lambda)$ 可以写成

$$p(\lambda)=\psi_m(\lambda)q(\lambda)+r(\lambda)$$

其中 $r(\lambda)$ 的次数小于 $\psi_m(\lambda)$ 的次数。

假如 $p(\lambda)$ 不能用 $\psi_m(\lambda)$ 整除，即 $r(\lambda)\neq0$，因为若 $p(\lambda)$ 是化零多项式，所以 $p(A)=0$，而且 $\psi_m(A)=0$，于是

$$r(A)=0$$

成立。而 $r(\lambda)$ 的次数比 $\psi_m(\lambda)$ 的次数小，此与 $\psi_m(\lambda)$ 是最小多项式矛盾。 ▮

推论 9.1.1 矩阵 A 的最小多项式是唯一的。

〔证明〕 设 ψ_m^1、ψ_m^2 是 A 的两个最小多项式，故其次数相等。由定理 9.1.5，ψ_m^1 和 ψ_m^2 可以互相整除，于是，有

$$\psi_m^1=a\psi_m^2, \qquad a\text{ 为纯量}$$

因 ψ_m^1，ψ_m^2 的首项（即最高次项）系数均为 1，故必有 $a=1$。 ▮

由定理 9.1.3，有

$$p(A)=0\Leftrightarrow p(PAP^{-1})=0$$

这表明：相似矩阵的最小多项式相同。这样，求 A 的最小多项式 $\psi_m(\lambda)$ 的问题可以归结为研究与 A 相似的 Jordan 标准形的最小多项式。

我们知道，Jordan 标准形的特征多项式是

$$D(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_r)^{m_r}$$

其中 $\lambda_1,\lambda_2,\cdots,\lambda_r$ 是 A 的相异特征值。可以证明：设 d_i 是 A 的 Jordan 形 J 中包含 λ_i 的最大 Jordan 块的阶数，则 A 的最小多项式为

$$\psi_m(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_r)^{d_r}$$

根据上面的结论，我们可以用初等变换求最小多项式。

例 9.1.2 求矩阵

$$A=\begin{pmatrix}3 & 0 & 8\\ 3 & -1 & 6\\ -2 & 0 & -5\end{pmatrix}$$

的最小多项式。

〔解〕 因 A 的 Jordan 标准形为

$$\begin{bmatrix} 1 & & \\ & \lambda+1 & \\ & & (\lambda+1)^2 \end{bmatrix}$$

所以最小多项式为

$$\psi_m=(\lambda+1)^2$$

下面我们再介绍一个求最小多项式的方法。

令 $B(\lambda)$ 为特征矩阵 $\lambda I-A$ 的伴随矩阵，$D(\lambda)=|\lambda I-A|$，$d(\lambda)$ 为 $B(\lambda)$ 中各元素的最高公因式。

定理 9.1.6 矩阵 A 的最小多项式为

$$\psi_m(\lambda)=\frac{D(\lambda)}{d(\lambda)} \tag{9-8}$$

〔证明〕 由 $B(\lambda)$ 的定义，有

$$B(\lambda)(\lambda I-A)=D(\lambda)I \tag{9-9}$$

令

$$C(\lambda)=\frac{B(\lambda)}{d(\lambda)} \tag{9-10}$$

因为矩阵 $B(\lambda)$ 的元素是 λ 的 $n-1$ 次多项式，所以可设 $C(\lambda)$ 中的元素是 λ 的 $m-1$ 次多项式，显然 $m<n$，于是 $C(\lambda)$ 可以写成

$$C(\lambda)=C_0\lambda^{m-1}+C_1\lambda^{m-2}+\cdots+C_{m-2}\lambda+C_{m-1} \tag{9-11}$$

其中 $C_0,C_1,\cdots,C_{m-1}$ 均为 n 阶常数矩阵

设

$$h(\lambda)=\frac{D(\lambda)}{d(\lambda)} \tag{9-12}$$

由式（9-9）可知，$D(\lambda)$ 能被 $d(\lambda)$ 整除，且应为 m 次多项式，故可设

$$h(\lambda)=\lambda^m+a_1\lambda^{m-1}+\cdots+a_{m-1}\lambda+a_m \tag{9-13}$$

先来证明 $h(\lambda)$ 是 A 的化零多项式。由式（9-9）与式（9-10），得

$$C(\lambda)(\lambda I-\mathrm{A})=\frac{B(\lambda)}{d(\lambda)}(\lambda I-A)$$

$$=\frac{D(\lambda)}{d(\lambda)}I=h(\lambda)I$$

将式（9-11）和式（9-13）代入上式，得

$$(C_0\lambda^{m-1}+C_1\lambda^{m-2}+\cdots+C_{m-2}\lambda+C_{m-1})(\lambda I-A)$$

$$=(\lambda^m+a_1\lambda^{m-1}+\cdots+a_{m-1}\lambda+a_m)I$$

比较上式两端的系数，有

$$\begin{cases} C_0 = = I \\ C_1 - C_0 A = a_1 I \\ C_2 - C_1 A = a_2 I \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ C_{m-1} - C_{m-2} A = a_{m-1} I \\ \qquad - C_{m-1} A = a_m I \end{cases}$$

以 $A^m, A^{m-1}, \cdots, A$ 依次右乘上述 $m+1$ 个等式，然后相加，得

$$A^m + a_1 A^{m-1} + \cdots + a_{m-1} A + a_m I = 0$$

即 $$h(A) = 0$$

其次，由多项式理论知，$\lambda - \mu \mid \psi_m(\lambda) - \psi_m(\mu)$，设

$$\psi_m(\lambda) - \psi_m(\mu) = (\lambda - \mu) g(\lambda, \mu)$$

以 λI 替换 λ，A 替换 μ，有

$$\psi_m(\lambda I) - \psi_m(A) = (\lambda I - A) g(\lambda I, A)$$

因为 $\psi_m(A) = 0$，所以，有

$$\psi_m(\lambda I) = (\lambda I - A) g(\lambda I, A)$$

或 $$\psi_m(\lambda) I = (\lambda I - A) g(\lambda I, A) \tag{9-14}$$

以 $B(\lambda)$ 左乘上式两端，有

$$\psi_m(\lambda) B(\lambda) = B(\lambda)(\lambda I - A) g(\lambda I, A)$$

由式（9-9），有

$$\psi_m(\lambda) B(\lambda) = D(\lambda) I g(\lambda I, A)$$

由式（9-10），有

$$\psi_m(\lambda) C(\lambda) = \frac{D(\lambda)}{d(\lambda)} g(\lambda I, A)$$

$$= h(\lambda) g(\lambda I, A)$$

于是 $h(\lambda)$ 能整除 $\psi_m(\lambda) C(\lambda)$ 中的每个元素，因 $C(\lambda)$ 中元素互质，故 $h(\lambda)$ 能整除 $\psi_m(\lambda)$，因而有

$$h(\lambda) \equiv \psi_m(\lambda)$$]

例 9.1.3 求矩阵

$$A = \begin{pmatrix} 2 & 1 & 0 \\ -4 & -2 & 0 \\ 2 & 1 & 0 \end{pmatrix}$$

的最小多项式 $\psi_m(A)$。

〔解〕 A 的特征多项式为

$$D(\lambda)=|\lambda I-A|=\lambda^3$$

而

$$B(\lambda)=\begin{pmatrix}\lambda^2+2\lambda & \lambda & 0\\ -4\lambda & \lambda^2-2\lambda & 0\\ 2\lambda & \lambda & \lambda^2\end{pmatrix}$$

于是 $d(\lambda)=\lambda$，因而

$$\psi_m(\lambda)=\frac{D(\lambda)}{d(\lambda)}=\lambda^2$$

矩阵的特征多项式与最小多项式还有下面的关系。

定理9.1.7 矩阵 A 的特征多项式 $D(\lambda)$ 的既约因式是 A 的最小多项式 $\psi_m(\lambda)$ 的因式，因此，特征多项式的根也是最小多项式的根。

〔证明〕 由式（9-14），有

$$(\psi_m(\lambda))^n=D(\lambda)\det(g(\lambda I,A))$$

于是 $D(\lambda)$ 的既约因式能整除 $(\psi_m(\lambda))^n$，因此也能整除 $\psi_m(\lambda)$。 ∎

推论9.1.2 若矩阵 A 的特征多项式 $D(\lambda)$ 没有重因式，则 A 的最小多项式 $\psi_m(\lambda)=D(\lambda)$。

矩阵的特征多项式和最小多项式是两种重要的又常见的化零多项式。

矩阵 A 的化零多项式在计算矩阵多项式 $p(A)$ 时十分有用，现简述如下：

设 n 阶方阵 A 的特征多项式和最小多项式分别为 $D(\lambda)=\det|\lambda I-A|$ 和 $\psi_m(\lambda)$，$p(\lambda)$ 为纯量 λ 的 s 次多项式，当 $s>n$（或 $s>m$）时，由多项式理论知

$$p(\lambda)=D(\lambda)q_1(\lambda)+r_1(\lambda)$$

或

$$p(\lambda)=\psi_m(\lambda)q_2(\lambda)+r_2(\lambda)$$

其中 $r_1(\lambda)$ 是次数小于 n 的多项式，$r_2(\lambda)$ 是次数小于 m 的多项式，因为 $D(\lambda)$、$\psi_m(\lambda)$ 是 A 的化零多项式，所以

$$p(A)=r_1(A) \quad 或 \quad p(A)=r_2(A)$$

这表明求 $p(A)$ 时可以用一个比 $p(\lambda)$ 次数低的多项式 $r_1(\lambda)$ 或 $r_2(\lambda)$ 来代替，而 $r_1(A)$ 或 $r_2(A)$ 的计算比 $p(A)$ 要简便。

例9.1.4 设

$$A=\begin{pmatrix}2 & 0 & 0\\ 1 & 1 & 1\\ 1 & -1 & 3\end{pmatrix}$$

$$p(\lambda)=\lambda^4-2\lambda^3+\lambda-1$$

试计算 $p(A)$。

〔解〕 A 的特征多项式 $D(\lambda)$ 和最小多项式 $\psi_m(\lambda)$ 分别为

$$D(\lambda)=(\lambda-2)^3, \qquad \psi_m(\lambda)=(\lambda-2)^2$$

由多项式除法得

$$p(\lambda)=D(\lambda)(\lambda+4)+(12\lambda^2-39\lambda+31)$$

或

$$p(\lambda)=\psi_m(\lambda)(\lambda^2+2\lambda+4)+(9\lambda-17)$$

于是

$$r_1(\lambda)=12\lambda^2-39\lambda+31, \qquad r_2(\lambda)=9\lambda-17$$

易知

$$p(A)=r_1(A)=r_2(A)=\begin{pmatrix}1 & 0 & 0\\ 9 & -8 & 9\\ 9 & -9 & 10\end{pmatrix}$$

最后，我们来讨论矩阵 A 的特征值、特征向量和 $p(A)$ 的特征值、特征向量之间的关系。

定理 9.1.8 设 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 n 阶矩阵 A 的 n 个特征值，$\boldsymbol{X}_i$ 是 A 关于 λ_i 的特征向量，$p(A)=a_m\lambda^m+a_{m-1}\lambda^{m-1}+\cdots+a_1\lambda+a_0$ 是纯量多项式，则

(1) $p(\lambda_1),p(\lambda_2),\cdots,p(\lambda_n)$ 是 $p(A)$ 的特征值，且 $p(A)$ 的特征值也只能是 $p(\lambda_1)$，$p(\lambda_2)\cdots,p(\lambda_n)$。

(2) $\boldsymbol{X}_i$ 是 $p(A)$ 关于 $p(\lambda_i)$ 的特征向量。

〔证明〕 由定理 9.1.3 的假设，有

$$A\boldsymbol{X}_i=\lambda_i\boldsymbol{X}_i$$

于是

$$\begin{aligned}p(A)\boldsymbol{X}_i&=(a_mA^m+a_{m-1}A^{m-1}+\cdots+a_1A+a_0I)\boldsymbol{X}_i\\&=a_mA^m\boldsymbol{X}_i+a_{m-1}A^{m-1}\boldsymbol{X}_i+\cdots+a_1A\boldsymbol{X}_i+a_0\boldsymbol{X}_i\\&=a_m\lambda_i^m\boldsymbol{X}_i+a_{m-1}\lambda_i^{m-1}\boldsymbol{X}_i+\cdots+a_1\lambda_i\boldsymbol{X}_i+a_0\boldsymbol{X}_i\\&=(a_m\lambda_i^m+a_{m-1}\lambda_i^{m-1}+\cdots+a_1\lambda_i+a_0)\boldsymbol{X}_i\\&=p(\lambda_i)\boldsymbol{X}_i\end{aligned}$$

上面的等式表明，$p(\lambda_i)$ 是 $p(A)$ 的特征值，且对应的特征向量为 $\boldsymbol{X}_i$。

下面证明 $p(A)$ 的特征值只能是 $p(\lambda_1)$，$p(\lambda_2)$，$\cdots$，$p(\lambda_n)$。

设 A 的 Jordan 标准形为

$$J=J_1\dotplus J_2\dotplus\cdots\dotplus J_r$$

因为 $p(A)=Pp(J)P^{-1}$，所以 $p(A)$ 与 $p(J)$ 相似，因而 $p(A)$ 与 $p(J)$ 有相同的特征值，$p(J_1),p(J_2),\cdots,p(J_r)$ 特征值的全体是 $p(J)$ 的特征值。由式（9-5），有

$$p(J_i)=\begin{pmatrix} p(\lambda_i) & & & * \\ & p(\lambda_i) & & \\ 0 & & \ddots & \\ & & & p(\lambda_i) \end{pmatrix}$$

可知 $p(J_i)$ 的特征值为 $p(\lambda_i)$。于是 $p(A)$的特征值只能是 $p(\lambda_1),p(\lambda_2),\cdots,p(\lambda_n)$。 ∎

需要指出的是，A 的特征向量也是 $p(A)$ 的特征向量，但 $p(A)$ 的特征向量则不一定是 A 的特征向量。譬如，设

$$A=\begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

A 的特征多项式为

$$\det(\lambda I-A)=(\lambda-1)^2(\lambda-2)$$

对应于 $\lambda=1$ 的 A 的特征向量为 (1,2,−1)，对应于 $\lambda=2$ 的 A 的特征向量为 (0,0,1)。

设 $p(\lambda)=\lambda^2-2\lambda+6$，则

$$p(A)=\begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ -1 & 1 & 6 \end{pmatrix}$$

$p(A)$ 的特征多项式为

$$\det(\lambda I-p(A))=(\lambda-5)^2(\lambda-6)$$
$$=(\lambda-p(1))^2(\lambda-p(2))$$

对应于 $\lambda=6$ 的 $p(A)$ 的特征向量是 (0,0,1)，对应于 $\lambda=5$ 的 $p(A)$ 的特征向量是 (1,2,−1)，(1,−1,2)。

显然，(1,−1,2) 不是 A 的对应于 $\lambda=1$ 的特征向量。

§9.2 矩阵谱上的函数

由例 9.1.2 可知，几个次数不同的多项式 $p(\lambda)$，$r_1(\lambda)$，$r_2(\lambda)$ 在满足某些条件时就有 $p(A)=r_1(A)=r_2(A)$。这就是本节要着重研究的问题。

定义 9.2.1 设 $A\in C^{n\times n}$，λ_1，λ_2，…，λ_r 是 A 的互异的特征值，A 的最小多项式为

$$\psi_m(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_r)^{d_r}$$

若一个给定函数 $f(\lambda)$ 具有足够多阶导数，下列 m 个值

$$f(\lambda_1), f'(\lambda_1), \cdots, f^{(d_1-1)}(\lambda_1)$$

$$f(\lambda_2), f'(\lambda_2), \cdots, f^{(d_2-1)}(\lambda_2)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$f(\lambda_r), f'(\lambda_r), \cdots, f^{(d_r-1)}(\lambda_r)$$

称为 $f(\lambda)$ 关于矩阵 A 的影谱上的值。如果这些值均存在，便称函数 $f(\lambda)$ 在 A 的谱上有定义。

显然，域 C 上的每一个多项式在任一矩阵 $A\in C^{n\times n}$ 的谱上有定义。

定理 9.2.1 矩阵 $A\in C^{n\times n}$ 的最小多项式 $\psi_m(\lambda)$ 在 A 的谱上的值均为零。

〔证明〕 显然

$$\psi_m(\lambda_1)=\psi_m(\lambda_2)=\cdots=\psi_m(\lambda_r)=0$$

$\psi_m(\lambda)$ 的一阶导数是

$$\begin{aligned}\psi_m'(\lambda)=&d_1(\lambda-\lambda_1)^{d_1-1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_r)^{d_r}\\&+d_2(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2-1}\cdots(\lambda-\lambda_r)^{d_r}\\&+\cdots\\&+d_r(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_r)^{d_r-1}\end{aligned}$$

于是有

$$\psi_m'(\lambda_1)=\psi_m'(\lambda_2)=\cdots=\psi_m'(\lambda_r)=0$$

类似地，有

$$\psi_m''(\lambda_1)=\psi_m''(\lambda_2)=\cdots=\psi_m''(\lambda_r)=0$$

如此继续下去，便知命题为真。 】

定理 9.2.2 若多项式 $p(\lambda)$ 在 A 的谱上的值均为零，则 $p(\lambda)$ 必可被 A 的最小多项式整除。

〔证明〕 因为

$$p(\lambda_1)=p'(\lambda_1)=\cdots=p^{(d_1-1)}(\lambda_1)=0$$

所以多项式 $p(\lambda)$ 必含有因子 $(\lambda-\lambda_1)^{d_1}$。同理，$p(\lambda)$ 含有 $(\lambda-\lambda_2)^{d_2},\cdots,(\lambda-\lambda_r)^{d_r}$，故命题得证。 】

定理 9.2.3 设 $A\in C^{n\times n}$，$p_1(\lambda)$、$p_2(\lambda)$ 为两个纯量多项式，则 $p_1(A)=p_2(A)$ 的充要条件是 $p_1(\lambda)$ 和 $p_2(\lambda)$ 在 A 的谱上的值相同。

〔证明〕 必要性 设 $p_1(A)=p_2(A)$。命

$$h(\lambda)=p_1(\lambda)-p_2(\lambda)$$

则

$$h(A)=p_1(A)-p_2(A)=0$$

即多项式 $h(\lambda)$ 为 A 的化零多项式。故 $h(\lambda)$ 可以被 A 的最小多项式 $\psi_m(\lambda)$ 整除。令

$$h(\lambda)=q(\lambda)\psi_m(\lambda)$$

即
$$p_1(\lambda)-p_2(\lambda)=q(\lambda)\psi_m(\lambda)$$

由定理9.2.1知，$\psi_m(\lambda)$在A的谱上的值为零，于是有

$$p_1^{(k)}(\lambda_j)-p_2^{(k)}(\lambda_j)=0 \qquad (j=1,2\cdots,r;\ k=0,1,2\cdots,d_j-1)$$

即
$$p_1^{(k)}(\lambda_j)=p_2^{(k)}(\lambda_j) \qquad (j=1,2,\cdots,r;\ k=0,1,2,\cdots,d_j-1)$$

充分性　设$p_1(\lambda)$、$p_2(\lambda)$在A的谱上的值相同，命

$$h(\lambda)=p_1(\lambda)-p_2(\lambda)$$

则$h(\lambda)$在A的谱上的值均为零，由定理9.2.2知，$h(\lambda)$可以被A的最小多项式$\psi_m(\lambda)$整除。设

$$p_1(\lambda)-p_2(\lambda)=h(\lambda)=q(\lambda)\psi_m(\lambda)$$

于是
$$p_1(A)-p_2(A)=h(A)=q(A)\varphi_m(A)=0$$

即
$$p_1(A)=p_2(A)$$
▮

推论 9.2.1　多项式$p(\lambda)$是A的化零多项式的充要条件是$p(\lambda)$在A的谱上的值为零。

下一节我们将利用矩阵多项式的性质，对一般函数$f(\lambda)$给出$f(A)$的定义。

§9.3　矩阵函数的定义

矩阵函数有多种定义方法，下面我们用矩阵多项式来给出一般的矩阵函数的概念。

设$A\in C^{n\times n}$，A的最小多项式为

$$\psi_m(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_r)^{d_r}$$

其中$\lambda_1,\lambda_2,\cdots,\lambda_r$为$A$的相异特征值。

由定理9.2.3知，在A的谱上一致的两个纯量多项式，其矩阵多项式相同。也就是说，给定A，任意矩阵多项式$p(A)$仅由$p(\lambda)$在A的谱上的值确定。

设$f(\lambda)$为任给函数，欲使矩阵函数$f(A)$是矩阵多项式$p(A)$的自然推广，就必须要求$f(\lambda)$在A的谱上有定义。为此

定义 9.3.1　设函数$f(\lambda)$在A的谱上有意义，即

$$f^{(k)}(\lambda_j) \qquad (j=1,2,\cdots,r;\ k=0,1,2,\cdots,d_j-1)$$

有确定的值。$p(\lambda)$为任一多项式，如果

$$f^{(k)}(\lambda_j)=p^{(k)}(\lambda_j)$$

即$f(\lambda)$与$p(\lambda)$在A的谱上相同，则矩阵函数$f(A)$定义为

$$f(A)\triangleq p(A)$$

由此定义可见，对于任给矩阵 A 和函数 $f(\lambda)$，$f(A)$ 不一定有定义。例如，若 $f(\lambda)=\ln\lambda$，那么当 A 的特征值为零或负数时，$f(\lambda_i)$ 没有意义，从而 $f(A)$ 就没有定义。此外，满足定义的 $p(\lambda)$ 不唯一。

§9.4 矩阵函数的性质

由于矩阵函数 $f(A)$ 就其实质而言，是一个矩阵多项式，因此它的性质容易研究。

利用矩阵函数的性质，不但可在已知 A 的 Jordan 标准形的情况下轻易地计算出 $f(A)$，还可以建立 A 和 $f(A)$ 的 Jordan 标准形之间的重要关系。

定理 9.4.1 设 $A\in C^{n\times n}$，则

(1) $$A\cdot f(A)=f(A)\cdot A \tag{9-15}$$

(2) $$(f_1+f_2)(A)=f_1(A)+f_2(A) \tag{9-16}$$

(3) $$(f_1f_2)(A)=f_1(A)\cdot f_2(A) \tag{9-17}$$

〔证明〕 若 f、f_1、f_2 是多项式，则式(9-15)、(9-16)和(9-17)显然成立。当 f、f_1、f_2 是解析函数时，由定义 9.3.1，$f(A)$ 等于矩阵多项式 $p(A)$，故这些性质也成立。 】

定理 9.4.2 设 $A\in C^{n\times n}$ 是分块对角矩阵

$$A=\begin{pmatrix}A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_t\end{pmatrix}$$

函数 $f(\lambda)$ 在 A 的谱上有意义，则

$$f(A)=\begin{pmatrix}f(A_1) & & & \\ & f(A_2) & & \\ & & \ddots & \\ & & & f(A_t)\end{pmatrix} \tag{9-18}$$

〔证明〕 首先，对任何多项式 $g(\lambda)$，有

$$g(A)=\begin{pmatrix}g(A_1) & & & \\ & g(A_2) & & \\ & & \ddots & \\ & & & g(A_t)\end{pmatrix}$$

于是，如果 $p(\lambda)$ 是在 A 的谱上与 $f(\lambda)$ 一致的多项式，则有

$$f(A)=p(A)=\begin{pmatrix}p(A_1) & & & \\ & p(A_2) & & \\ & & \ddots & \\ & & & p(A_t)\end{pmatrix}$$

其次，由于 $A_j(1\leqslant j\leqslant t)$ 的谱是 A 的谱的子集，$f(\lambda)$ 在每个 $A_j(1\leqslant j\leqslant t)$ 的谱上有定义，而且，因为假定 $f(\lambda)$ 和 $p(\lambda)$ 在 A 的谱上有相同的值，所以它们在 $A_j(1\leqslant j\leqslant t)$ 的谱上也有相同的值。于是 $f(A_j)=p(A_j)$，故式（9-18）成立。 ▮

定理 9.4.3 设 A，B，$P\in C^{n\times n}$，其中 $B=P^{-1}AP$，且 $f(\lambda)$ 在 A 的谱上有意义，则

$$f(B)=P^{-1}f(A)P$$

〔证明〕 因为 A、B 相似，故它们有相同的最小多项式。于是，若 $p(\lambda)$ 是在 A 的谱上与 $f(\lambda)$ 一致的多项式，则它也是在 B 的谱上与 $f(\lambda)$ 一致的多项式。从而，我们有

$$f(A)=p(A),\quad f(B)=p(B),\quad p(B)=P^{-1}p(A)P$$

于是

$$f(B)=P^{-1}f(A)P$$

▮

由定理 9.4.2 和定理 9.4.3 可以直接得出下面的定理。

定理 9.4.4 设 $A\in C^{n\times n}$，$A=PJP^{-1}$，其中

$$J=\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_t \end{pmatrix}$$

是 J 的 Jordan 标准形，J_i 是对应于 λ_i 的 Jordan 块，则

$$f(A)=P\begin{pmatrix} f(J_1) & & & \\ & f(J_2) & & \\ & & \ddots & \\ & & & f(J_t) \end{pmatrix}P^{-1} \tag{9-19}$$

定理 9.4.5 设 J_i 是对应于 λ_i 的 Jordan 块，

$$J_i=\begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{l\times l}$$

如果 $f(\lambda)$ 在 λ_i 的邻域内具有 $l-1$ 阶导数，则

$$f(J_i)=\begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{(l-1)!}f^{(l-1)}(\lambda_i) \\ & f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{1}{(l-2)!}f^{(l-2)}(\lambda_i) \\ & & \ddots & & \\ & & & \ddots & \\ & & & & f(\lambda_i) \end{pmatrix} \tag{9-20}$$

〔证明〕 J_i 的最小多项式为

$$(\lambda-\lambda_i)^l$$

$f(\lambda)$ 在 J_i 的谱上的值是

$$f(\lambda_i),f'(\lambda_i),\cdots,f^{(l-1)}(\lambda_i)$$

令
$$p(\lambda)=f(\lambda_i)+f'(\lambda_i)(\lambda-\lambda_i)+\cdots+\frac{f^{(l-1)}(\lambda_i)}{(l-1)!}(\lambda-\lambda_i)^{l-1}$$

显然 $f(\lambda)$ 与 $p(\lambda)$ 在 J_i 的谱上的值相同，所以 $f(J_i)=p(J_i)$，于是

$$f(J_i)=f(\lambda_i)I+f'(\lambda_i)(J_i-\lambda_iI)+\cdots+\frac{f^{(l-1)}(\lambda_i)}{(l-1)!}(J_i-\lambda_iI)^{l-1} \tag{9-21}$$

计算 $J_i-\lambda_iI$ 的幂，有

$$(J_i-\lambda_iI)^j=\begin{pmatrix}0 & 1 & 0 & \cdots & 0 & 0\\ & 0 & 1 & \cdots & \cdots & 0\\ & & 0 & 1 & \cdots & 0\\ & & & \ddots & \ddots & \vdots\\ & & & & \ddots & 1\\ & & & & & 0\end{pmatrix}^j=\begin{pmatrix}0 & \cdots & 0 & \overset{(j+1)\text{列}}{1} & 0 & \cdots & 0\\ & \ddots & & & \ddots & & \vdots\\ & & & & & & 1\\ & & & \ddots & & & \vdots\\ & & & & & & 0\end{pmatrix}$$

代入式（9-21），便得到式（9-20）。 ∎

式（9-19）称为矩阵函数 f(A) 的 Jordan 表示。

例 9.4.1 设矩阵

$$A=\begin{pmatrix}2 & 0 & 0\\ 1 & 1 & 1\\ 1 & -1 & 3\end{pmatrix}$$

求 $f(A)$ 的 Jordan 表示，并计算 e^A 和 $\sin A$。

〔解〕 矩阵 A 的 Jordan 标准形为

$$J=\begin{pmatrix}2 & 1 & 0\\ 0 & 2 & 0\\ 0 & 0 & 2\end{pmatrix}$$

变换矩阵 P 为

$$P=\begin{pmatrix}0 & 1 & 1\\ 1 & 0 & 0\\ 1 & 0 & -1\end{pmatrix},\quad P^{-1}=\begin{pmatrix}0 & 1 & 0\\ 1 & -1 & 1\\ 0 & 1 & -1\end{pmatrix}$$

而
$$f(J_1)=\begin{pmatrix} f(2) & f'(2) \\ 0 & f(2) \end{pmatrix},\quad f(J_2)=(f(2))$$

于是，矩阵函数 $f(A)$ 的 Jordan 表示为

$$f(A)=P\begin{pmatrix} f(2) & f'(2) & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(2) \end{pmatrix}P^{-1}$$

$$=\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix}\begin{pmatrix} f(2) & f'(2) & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(2) \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

$$=\begin{pmatrix} f(2) & 0 & 0 \\ f'(2) & f(2)-f'(2) & f'(2) \\ f'(2) & -f'(2) & f'(2)+f(2) \end{pmatrix} \tag{9-22}$$

当 $f(x)=e^x$ 时，由式（9-22）得

$$e^A=\begin{pmatrix} e^2 & 0 & 0 \\ e^2 & 0 & e^2 \\ e^2 & -e^2 & 2e^2 \end{pmatrix}$$

当 $f(x)=\sin x$ 时，由式（9-22）得

$$\sin A=\begin{pmatrix} \sin 2 & 0 & 0 \\ \cos 2 & \sin 2-\cos 2 & \cos 2 \\ \cos 2 & -\cos 2 & \sin 2+\cos 2 \end{pmatrix}$$

若 $f(x)=e^{tx}$，其中 t 是参数，$t\in(-\infty,\infty)$，则由式（9-20）可得

$$f(J_i)=\begin{pmatrix} e^{\lambda_i t} & te^{\lambda_i t} & \frac{t^2}{2!}e^{\lambda_i t} & \cdots & \frac{t^{d_i-1}}{(d_i-1)!}e^{\lambda_i t} \\ & e^{\lambda_i t} & te^{\lambda_i t} & \ddots & \vdots \\ & & e^{\lambda_i t} & \ddots & \\ & & & \ddots & te^{\lambda_i t} \\ & & & & e^{\lambda_i t} \end{pmatrix}_{d_i\times d_i}$$

$$
=e^{\lambda_i t}\begin{pmatrix}
1 & t & \frac{1}{2!}t^2 & \cdots & \frac{t^{d_i-1}}{(d_i-1)!} \\
 & 1 & \ddots & \ddots & \vdots \\
 & & \ddots & \ddots & \vdots \\
 & & & \ddots & t \\
 & & & & 1
\end{pmatrix}_{d_i\times d_i}
$$

§9.5 矩阵函数的 Lagrange-Sylvester 内插多项式表示

根据矩阵函数的定义知，可以利用和纯量函数 $f(\lambda)$ 在矩阵 A 的谱上一致的多项式 $p(\lambda)$ 确定 $f(A)$。一般说来，这种多项式不是唯一的，而有无数个。但是其中比 A 的最小多项式 $\psi_m(\lambda)$ 次数小的只有 Lagrange-Sylvester 内插多项式，现在介绍如下。

设 n 阶矩阵 A 的最小多项式为

$$
\psi_m(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_r)^{d_r}
$$

$$
(d_1+d_2+\cdots+d_r=m)
$$

设多项式 $p(\lambda)$ 的次数小于 m。则可将 $p(\lambda)/\psi_m(\lambda)$ 展开成部分分式

$$
\frac{p(\lambda)}{\psi_m(\lambda)}=\sum_{k=1}^{r}\left[\frac{a_{k1}}{(\lambda-\lambda_k)^{d_k}}+\frac{a_{k2}}{(\lambda-\lambda_k)^{d_k-1}}+\cdots+\frac{a_{kd_k}}{(\lambda-\lambda_k)}\right] \tag{9-23}
$$

不难验证上式中系数 a_{kl} 可由

$$
a_{kl}=\frac{1}{(l-1)!}\left[\frac{\mathrm{d}^{l-1}}{\mathrm{d}\lambda^{l-1}}\left(\frac{(\lambda-\lambda_k)^{d_k}p(\lambda)}{\psi_m(\lambda)}\right)\right]_{\lambda=\lambda_k}
$$

$$
(k=1,2,\cdots,r;\ \ l=1,2,\cdots,d_k) \tag{9-24}
$$

确定。式（9-24）中 $p(\lambda)$ 的数值实质上是 $p(\lambda)$ 在 A 的影谱上的值，欲使 $p(A)=f(A)$ 需要 $p(\lambda)$ 在 A 的影谱上的值与 $f(\lambda)$ 在 A 的影谱上的值相同，故式（9-24）可改为

$$
a_{kl}=\frac{1}{(l-1)!}\left[\frac{\mathrm{d}^{l-1}}{\mathrm{d}\lambda^{l-1}}\left(\frac{(\lambda-\lambda_k)^{d_k}f(\lambda)}{\psi_m(\lambda)}\right)\right]_{\lambda=\lambda_k} \tag{9-25}
$$

由式（9-23）得

$$
p(\lambda)=\sum_{k=1}^{r}(a_{k1}+a_{k2}(\lambda-\lambda_k)+\cdots+a_{kd_k}(\lambda-\lambda_k)^{d_k-1})\varphi_k(\lambda) \tag{9-26}
$$

式中
$$\varphi_k(\lambda)=\frac{\psi_m(\lambda)}{(\lambda-\lambda_k)^{d_k}}=(\lambda-\lambda_1)^{d_1}\cdots(\lambda-\lambda_{k-1})^{d_{k-1}}(\lambda-\lambda_{k+1})^{d_{k+1}}\cdots(\lambda-\lambda_r)^{d_r} \quad (9\text{-}27)$$

由式（9-26）得到的多项式 $p(\lambda)$ 便是 Lagrange-Sylvester 内插多项式，它的次数为 $m-1$。$p(\lambda)$ 与 $f(\lambda)$ 在 A 的影谱上的值一致，于是

$$f(A)=\sum_{k=1}^{r}[a_{k1}I+a_{k2}(A-\lambda_k I)+\cdots+a_{kd_k}(A-\lambda_k I)^{d_k-1}]\varphi_k(A) \quad (9\text{-}28)$$

式（9-28）称为矩阵函数 $f(A)$ 的 Lagrange-Sylvester 内插多项式表示。

例 9.5.1 设

$$A=\begin{pmatrix}-1 & 1 & 0\\ -4 & 3 & 0\\ 1 & 0 & 2\end{pmatrix}$$

试求 $f(A)$ 的 Lagrange-Sylvester 内插多项式表示，并求 e^A。

〔解〕 矩阵 A 的最小多项式为

$$\psi_m(\lambda)=(\lambda-1)^2(\lambda-2)$$

于是 $\lambda_1=1$，$\lambda_2=2$，$d_1=2$，$d_2=1$。代入式（9-25）得

$$a_{11}=\left[\frac{f(\lambda_1)}{\varphi(\lambda_1)}\right]=-f(1)$$

$$a_{12}=\frac{d}{d\lambda}\left[\frac{f(\lambda)}{\varphi_1(\lambda)}\right]_{\lambda=\lambda_1}=\frac{d}{d\lambda}\left[\frac{f(\lambda)}{\lambda-2}\right]_{\lambda=1}$$

$$=-f'(1)-f(1)$$

$$a_{21}=\left[\frac{f(\lambda)}{\varphi_2(\lambda)}\right]_{\lambda=\lambda_2}=\left[\frac{f(\lambda)}{\lambda-1}\right]_{\lambda=2}=f(2)$$

代入式（9-28）有

$$f(A)=\{-f(1)I-(f'(1)+f(1))(A-I)\}(A-2I)+f(2)(A-I)^2 \quad (9\text{-}29)$$

若 $f(\lambda)=e^\lambda$，则有

$$f(1)=e\ ,\ f'(1)=e,\ f(2)=e^2$$

于是，由式（9-29）有

$$e^A=[-eI-(e+e)(A-I)](A-2I)+e^2(A-I)^2$$

$$=(eI-2eA)(A-2I)+e^2(A-I)^2$$

$$=\begin{pmatrix} -e & e & 0 \\ -4e & 3e & 0 \\ 3e-e^2 & -2e+e^2 & e^2 \end{pmatrix}$$

在本节最后我们介绍矩阵函数的多项式表示。由于式（9-26）中的 $p(\lambda)$ 实质上是一个 $m-1$ 次的多项式，所以将式（9-26）的右端按 λ 的降幂排列便得

$$p(\lambda)=a_{m-1}\lambda^{m-1}+a_{m-2}\lambda^{m-2}+\cdots+a_1\lambda+a_0$$

若能求得 $a_j(j=0,1,2,\cdots,m-1)$，则矩阵函数 $f(A)$ 为

$$f(A)=a_{m-1}A^{m-1}+a_{m-2}A^{m-2}+\cdots+a_1A+a_0I \qquad (9\text{-}30)$$

式（9-30）便是矩阵函数 $f(A)$ 的多项式表示。m 个系数 a_0，$a_1,\cdots,a_{m-1}$ 的确定可由 $p(\lambda)$ 与 $f(\lambda)$ 在 A 的影谱上的值一致而得到，即由

$$\begin{cases} p(\lambda_k)=f(\lambda_k) \\ p'(\lambda_k)=f'(\lambda_k) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ p^{(d_k-1)}(\lambda_k)=f^{(d_k-1)}(\lambda_k) \end{cases} \qquad (k=1,2,\cdots,r) \qquad (9\text{-}31)$$

可以解得 $a_0,a_1,\cdots,a_{m-1}$。

例 9.5.2 设

$$A=\begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

试求 $f(A)$ 的多项式表示，并求 e^A。

〔解〕 A 的最小多项式为

$$\psi_m(\lambda)=(\lambda-1)^2(\lambda-2)$$

所以 $$p(\lambda)=a_2\lambda^2+a_1\lambda+a_0$$

于是由式（9-31）可得

$$\begin{cases} a_2+a_1+a_0=e \\ 4a_2+2a_1+a_0=e^2 \\ 2a_2+a_1=e \end{cases}$$

解之得

$$a_0=e^2-2e,\quad a_1=5e-2e^2,\quad a_2=e^2-2e$$

所以

$$p(\lambda)=(e^2-2e)\lambda^2+(5e-2e^2)\lambda+(e^2-2e)$$

$$f(A)=(e^2-2e)A^2+(5e-2e^2)A+(e^2-2e)I$$

$$=\begin{pmatrix} -e & e & 0 \\ -4e & 3e & 0 \\ 3e-e^2 & -2e+e^2 & e^2 \end{pmatrix}$$

§9.6 矩阵函数的谱分解与矩阵分量

若将式（9-25）代入 Lagrange-Sylvester 内插多项式（9-26），把含有 $f(\lambda_k)$，$f'(\lambda_k),\cdots,f^{(d_k-1)}(\lambda_k)(k=1,2,\cdots,r)$ 的项各自分开，整理后可以写成

$$p(\lambda)=\sum_{k=1}^{r}[f(\lambda_k)p_{k0}(\lambda)+f'(\lambda_k)p_{k1}(\lambda)+\cdots$$
$$+f^{(d_k-1)}(\lambda_k)p_{k,d_k-1}(\lambda)] \qquad (9\text{-}32)$$

其中 $p_{kl}(\lambda)(k=1,2,\cdots,r;\ l=0,1,2,\cdots,d_k-1)$ 是次数小于 m 的多项式，它完全由 A 的最小多项式 $\psi_m(\lambda)$ 确定，与所取的函数 $f(\lambda)$ 无关。根据 $p(\lambda)$ 与 $f(\lambda)$ 在 A 的影谱上的值相同，不难验证 $p_{kl}(\lambda)$ 在 A 的影谱上的值为

$$\begin{cases} p_{kl}^{(l)}(\lambda_k)=1 \\ p_{kl}^{(i)}(\lambda_k)=0 & i\neq l \\ p_{kl}^{(i)}(\lambda_j)=0 & j\neq k \end{cases}$$

由此，每一个 $p_{kl}(\lambda)$ 实质上也是一个 Lagrange-Sylvester 内插多项式。当 $p_{kl}(\lambda)$ 确定后，便可由式（9-32）得

$$f(A)=\sum_{k=1}^{r}[f(\lambda_k)p_{k0}(A)+f'(\lambda_k)p_{k1}(A)+\cdots$$
$$+f^{(d_k-1)}(\lambda_k)p_{k,d_k-1}(A)] \qquad (9\text{-}33)$$

其中 $p_{kl}(A)$ $(k=1,2,\cdots,r;\ l=0,1,2,\cdots,d_k-1)$ 与 $f(\lambda)$ 完全无关，是仅由 A 决定的常数矩阵，称为矩阵 A 的分量。式（9-33）常称为矩阵函数的基本公式。由于 $f(\lambda_k)$，$f'(\lambda_k),\cdots,f^{(d_k-1)}(\lambda_k)$ 是 $f(A)$ 的影谱上的值，所以式（9-33）也可称为谱分解。

关于矩阵分量我们有下面重要的定理。

定理9.6.1 任意矩阵 $A\in C^{n\times n}$ 的分量 $p_{kl}(A)$ $(k=1,2,\cdots,r;\ l=0,1,2,\cdots,d_k-1)$ 是线性无关的。特别是，m 个矩阵分量均不是零矩阵。

〔证明〕 先来证明纯量多项式 $p_{kl}(\lambda)$ $(k=1,2,\cdots,r;\ l=0,1,2,\cdots,d_k-1)$ 是线性无关的。

假设存在常数 $c_{kl}\in C$，对 $\forall\lambda\in C$，有

$$\sum_{k=1}^{r}\sum_{l=0}^{d_k-1}c_{kl}p_{kl}(\lambda)=0$$

成立。这意味着满足

$$f^{(l)}(\lambda_k)=c_{kl} \qquad (k=1,2,\cdots,r;\ l=0,1,2,\cdots,d_k-1)$$

的函数 $f(\lambda)$ 的 Lagrange-Sylvester 内插多项式 $p(\lambda)$ 恒等于零。因此有

$$p^{(l)}(\lambda_k)=f^{(l)}(\lambda_k)=c_{kl}=0$$

即 $p_{kl}(\lambda)$ 是线性无关的。

但是 $p(\lambda)=\sum_{k=1}^{r}\sum_{l=0}^{d_k-1}c_{kl}p_{kl}(\lambda)$ 的次数小于最小多项式的次数，因 $p_{kl}(\lambda)$ 是线性无关的，故除了 $c_{kl}=0$ $(k=1,2,\cdots,r;\ l=0,1,2,\cdots,d_k-1)$ 之外，不会得到

$$p(A)=\sum_{k=1}^{r}\sum_{l=0}^{d_k-1}c_{kl}p_{kl}(A)=0$$

即 $p_{kl}(A)$ 是线性无关的。]

下面我们来介绍一种比较实用的求矩阵分量的方法。

因为 $p_{kl}(A)$ 与 f 无关而只与 A 有关，所以在确定 $p_{kl}(A)$ 时可以选择最方便的函数。考虑函数

$$f(z)=\frac{1}{\lambda-z}$$

则
$$f(A)=(\lambda I-A)^{-1}=\frac{(\lambda I-A)^*}{D(\lambda)} \tag{9-34}$$

其中 $(\lambda I-A)^*$ 为 $\lambda I-A$ 的伴随矩阵，$D(\lambda)$ 为 A 的特征多项式。式(9-34)可以改写成

$$\frac{(\lambda I-A)^*}{D(\lambda)}=\frac{R(\lambda)}{\psi_m(\lambda)} \tag{9-35}$$

其中 $\psi_m(\lambda)$ 为 A 的最小多项式，$R(\lambda)$ 是以 λ 的多项式（次数均比 $\psi_m(\lambda)$ 的小）为元素的矩阵。将式(9-35)的右端展成部分分式

$$\frac{R(\lambda)}{\psi_m(\lambda)}=\sum_{k=1}^{r}\left(\frac{R_{k0}}{\lambda-\lambda_k}+\frac{R_{k1}}{(\lambda-\lambda_k)^2}+\cdots+\frac{R_{kd_k-1}}{(\lambda-\lambda_k)^{d_k}}\right) \tag{9-36}$$

其中 R_{kl} 是常数矩阵，可以用和纯量部分分式展开公式相同的形式给出：

$$R_{kl}=\frac{1}{(d_k-l-1)!}\left[\frac{\mathrm{d}^{(d_k-l-1)}}{\mathrm{d}\lambda^{(d_k-l-1)}}\left(\frac{(\lambda-\lambda_k)^{d_k}R(\lambda)}{\psi_m(\lambda)}\right)\right]_{\lambda=\lambda_k} \tag{9-37}$$

此外，对于 $f(z)=1/(\lambda-z)$，因为

$$f^{(l)}(\lambda_k)=\frac{l!}{(\lambda-\lambda_k)^l}$$

于是 $f(A)$ 可以表示为

$$f(A)=\sum_{k=1}^{r}\left[\frac{1}{\lambda-\lambda_k}p_{k0}(A)+\frac{1}{(\lambda-\lambda_k)^2}p_{k1}(A)+\cdots\right.$$

$$\left.+\frac{d_{k-1}!}{(\lambda-\lambda_k)^{d_k}}p_{kd_k-1}(A)\right] \qquad (9\text{-}38)$$

由式（9-36）和式（9-38）可以给出 A 的分量：

$$p_{kl}(A)=\frac{1}{l!}R_{kl} \qquad (k=1,2,\cdots,r;\ l=0,1,2,\cdots,d_k-1)$$

R_{kl} 可以由式（9-37）计算。

例 9.6.1 设

$$A=\begin{pmatrix}2 & 0 & 0\\ 1 & 1 & 1\\ 1 & -1 & 3\end{pmatrix}$$

求 A 的分量，并计算 e^A，$\sin A$。

〔解〕 A 的特征多项式为

$$|\lambda I-A|=(\lambda-2)^3$$

于是 A 的最小多项式为

$$\psi_m(\lambda)=(\lambda-2)^2$$

矩阵 $\lambda I-A$ 的伴随矩阵是

$$(\lambda I-A)^*=\begin{pmatrix}(\lambda-2)^2 & 0 & 0\\ \lambda-2 & (\lambda-2)(\lambda-3) & \lambda+2\\ \lambda-2 & -\lambda+2 & (\lambda-2)(\lambda-1)\end{pmatrix}$$

由式（9-35），有

$$R(\lambda)=\begin{pmatrix}\lambda-2 & 0 & 0\\ 1 & \lambda+3 & 1\\ 1 & -1 & \lambda-1\end{pmatrix}$$

再由式（9-37），有

$$R_{10}=\left[\frac{d}{d\lambda}\ \frac{(\lambda-2)^2R(\lambda)}{(\lambda-2)^2}\right]_{\lambda=2}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$$

$$R_{11}=\left[\frac{(\lambda-2)^2R(\lambda)}{(\lambda-2)^2}\right]_{\lambda=2}=\begin{pmatrix}0 & 0 & 0\\ 1 & -1 & 1\\ 1 & -1 & 1\end{pmatrix}$$

代入式(9-33)，得

$$f(A)=f(2)R_{10}(A)+f'(2)R_{11}(A) \tag{9-39}$$

以 $f(\lambda)=e^{\lambda}$ 代入式(9-39)，有

$$e^A=e^2R_{10}(A)+e^2R_{11}(A)$$

$$=\begin{pmatrix}e^2 & 0 & 0\\ e^2 & 0 & e^2\\ e^2 & -e^2 & 2e^2\end{pmatrix}$$

以 $f(\lambda)=\sin\lambda$ 代入式(9-39)，有

$$\sin A=\sin 2R_{10}(A)+\cos 2R_{11}(A)$$

$$=\begin{pmatrix}\sin 2 & 0 & 0\\ \cos 2 & \sin 2-\cos 2 & \cos 2\\ \cos 2 & -\cos 2 & \sin 2+\cos 2\end{pmatrix}$$

§9.7 矩阵函数的幂级数表示

这一节我们来讨论矩阵函数的幂级数表示，并证明作为幂级数和的矩阵函数，在其幂级数收敛的情况下与§9.3给出的矩阵函数一致。

我们知道，对于任意 $\lambda\in C$，纯量级数

$$e^{\lambda}=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!},\quad \cos\lambda=\sum_{k=0}^{\infty}(-1)^k\frac{\lambda^{2k}}{(2k)!},\quad \sin\lambda=\sum_{k=0}^{\infty}(-1)^k\frac{\lambda^{2k+1}}{(2k+1)!}$$

是绝对收敛的。相应的分别把 e^A，$\cos A$ 和 $\sin A$ 定义为

$$e^A=\sum_{k=0}^{\infty}\frac{A^k}{k!},\quad \cos A=\sum_{k=0}^{\infty}(-1)^k\frac{A^{2k}}{(2k)!},\quad \sin A=\sum_{k=0}^{\infty}(-1)^k\frac{A^{2k+1}}{(2k+1)!}$$

下面的定理表明，作为绝对收敛幂级数和定义的矩阵函数与§9.3中所定义的矩阵函数是一致的。

定理 9.7.1 设 A 的最小多项式为

$$\psi_m(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_r)^{d_r}$$

$f(\lambda),f_n(\lambda)\ (n=1,2,\cdots)$ 都是包含 $\lambda_1,\lambda_2,\cdots,\lambda_r$ 的开域上的解析函数。当 $n\to\infty$，$f_n(\lambda)$ 在 A 的谱上收敛于 $f(\lambda)$ 时，即当

$$\lim_{n\to\infty} f_n^{(l)}(\lambda_k)=f^{(l)}(\lambda_k) \tag{9-40}$$

$(k=1,2,\cdots,r;\ l=0,1,2,\cdots,d_k-1)$ 时有

$$\lim_{n\to\infty} f_n(A)=f(A) \tag{9-41}$$

反之，若式（9-41）成立，则式（9-40）也成立。

〔证明〕 设式（9-40）成立。$f_n(A)$ 的矩阵分量表示为

$$f_n(A)=\sum_{k=1}^{r}(f_n(\lambda_k)p_{k0}(A)+f_n'(\lambda_k)p_{k1}(A)+\cdots$$

$$+f_n^{(d_k-1)}(\lambda_k)p_{k,d_k-1}(A)) \tag{9-42}$$

$f(A)$ 的矩阵分量表示为

$$f(A)=\sum_{k=1}^{r}(f(\lambda_k)p_{k0}(A)+f'(\lambda_k)p_{k1}(A)+\cdots$$

$$+f^{(d_k-1)}(\lambda_k)p_{k,d_k-1}(A)) \tag{9-43}$$

显然，若式（9-40）成立，则 $\lim\limits_{n\to\infty}f_n(A)$ 存在，且等于 $f(A)$。

反之，假定式（9-41）成立，即

$$\lim_{n\to\infty} f_n(A)-f(A)=0$$

由定理 9.6.1 知，A 的矩阵分量是线性无关的，比较式（9-42）和式（9-43）知，A 的矩阵分量的系数为零，即

$$\lim_{n\to\infty} f_n^{(l)}(\lambda_k)-f^{(l)}(\lambda_k)=0$$

$(k=1,2,\cdots,r;\ l=0,1,2,\cdots,d_k-1)$。 ∎

推论 9.7.1 设纯量级数 $\sum\limits_{k=0}^{\infty}c_k\lambda^k$ 的收敛半径为 $R>0$，这时对全部特征值都在半径为 R 的圆内的矩阵 $A\in C^{n\times n}(|\lambda_k|<R;\ k=1,2,\cdots,r)$ 矩阵级数

$$\sum_{k=0}^{\infty}c_kA^k$$

绝对收敛，而且为

$$f(\lambda)=\sum_{k=0}^{\infty}c_k\lambda^k$$

时，$f(A)$（按 §9.3 定义的矩阵函数）满足

$$f(A)=\sum_{k=0}^{\infty}c_kA^k$$

〔证明〕 设

$$f_n(\lambda)=\sum_{k=0}^{n}c_k\lambda^k$$

由于幂级数与其微分有相同的收敛半径，$f_n^{(1)}(\lambda)$ 对于 $|\lambda|<R$ 绝对收敛，因此，对于 $|\lambda_k|<R$ 的矩阵 A，由定理 9.7.1，有

$$\lim_{n\to\infty}f_n(A)=\sum_{k=0}^{\infty}c_kA^k=f(A)$$ ■

例 9.7.1 因为

$$(1-\lambda)^{-1}=1+\lambda+\lambda^2+\cdots+\lambda^k+\cdots \qquad (R<1)$$

故有 $$(I-A)^{-1}=I+A+A^2+\cdots+A^k+\cdots$$

$$(|\lambda_k|<1;\ k=1,2,\cdots,r)$$

又因 $$\log(1+\lambda)=\lambda-\frac{1}{2}\lambda^2+\frac{1}{3}\lambda^3-\cdots \qquad (R<1)$$

故有 $$\log(I+A)=A-\frac{1}{2}A^2+\frac{1}{3}A^3-\cdots$$

$$(|\lambda_k|<1;\ k=1,2,\cdots,r)$$

习　　题

9-1　设 A 为 n 阶矩阵，$f(A)$ 有定义（在定义 9.3.1 的意义下），试证 $f(A^T)$ 也有定义，且

$$f(A^T)=[f(A)]^T$$

9-2　若 A 为非奇异矩阵，$f(A)$ 有定义，那么 $f(A^{-1})$ 是否有定义？试举例说明。

9-3　设 A 为非奇异矩阵，$f(\lambda)=\lambda^{q/p}$，试证 $f(A)$ 有定义，且

$$[f(A)]^p=A^q$$

9-4　设

$$A=\begin{bmatrix}0 & -1\\ 4 & 4\end{bmatrix}$$

计算 e^A。

9-5　设

$$A=\begin{bmatrix}-3 & 2 & 1\\ 1 & -2 & 1\\ 1 & 2 & -2\end{bmatrix}$$

计算 $\operatorname{arctg}(A/4)$。

9-6 设

$$A=\begin{pmatrix}0 & 1 & 0\\0 & 0 & 1\\1 & 0 & 0\end{pmatrix}$$

(1) 计算 A 的矩阵分量。

(2) 计算 e^A, $\sin A$。

9-7 试证对任意矩阵 A 均有

$$\cos 2A=\cos^2 A-\sin^2 A$$

$$\sin^2 A=\frac{1}{2}(I-\cos 2A)$$

$$e^{A+2\pi i I}=e^A$$

9-8 试证 $\lim\limits_{t\to\infty} e^{At}=0$ 当且仅当 A 的所有特征值的实部均为负数。

9-9 证明推论 9.2.1。

9-10 设 k 为纯量，A 为任意矩阵，证明

$$e^{kA}=e^{kI}A$$

9-11 设 $A\in C^{n\times n}$ 的 Jordan 标准形 $J=T^{-1}AT$ 为

$$J=\begin{pmatrix}J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r\end{pmatrix}_{n\times n}$$

求矩阵函数 e^{At} 的 Jordan 标准形表示。

第十章　矩阵微分方程

这一章和下一章我们来讨论矩阵方程，本章先讨论矩阵微分方程，下一章讨论矩阵代数方程。

本章讨论的内容有，形如 $\mathrm{d}X(t)/\mathrm{d}t=A(t)X(t)$ 和 $\mathrm{d}X(t)/\mathrm{d}t=A(t)X(t)+X(t)B(t)$ 的矩阵微分方程，状态转移矩阵，线性齐次和非齐次向量微分方程。

§10.1　形如 $\mathrm{d}X(t)/\mathrm{d}t=A(t)X(t)$ 的方程

形如

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=A(t)X(t),\qquad X(t_0)=C \tag{10-1}$$

的方程称为矩阵微分方程，其中 $A(t)$ ）是定义在 (t_0,t_1) 上的分段连续 n 阶函数矩阵，$X(t)$ 是 n 阶未知函数矩阵，C 是 n 阶常数矩阵。

方程(10-1)实际上是一个含有 n^2 个未知函数的常微分方程组。

定理 10.1.1　矩阵微分方程(10-1)的解存在且唯一。

〔证明〕　作矩阵序列 $\{X_n(t)\}$：

$$\begin{cases} X_0(t)=C, \\ X_1(t)=C+\int_{t_0}^{t}A(s)X_0(s)\mathrm{d}s \\ X_2(t)=C+\int_{t}^{t}A(s)X_1(s)\mathrm{d}s \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ X_n(t)=C+\int_{t_0}^{t}A(s)X_{n-1}(s)\mathrm{d}s \\ X_{n+1}(t)=C+\int_{t_0}^{t}A(s)X_n(s)\mathrm{d}s \end{cases} \tag{10-2}$$

于是我们有

$$X_{n+1}(t)-X_n(t)=\int_{t_0}^{t}A(s)(X_n(s)-X_{n-1}(s))\mathrm{d}s$$

设 $M=\max\limits_{t_0\leqslant t\leqslant t_1}\|A(s)\|$, $\|A(s)\|=\sum\limits_{i,j=1}^{n}|a_{ij}(s)|$, 则

$$\|X_{n+1}(t)-X_n(t)\|=\left\|\int_{t_0}^{t}A(s)(X_n(s)-X_{n-1}(s)\mathrm{d}s\right\|$$

$$\leqslant M\int_{t_0}^{t}\|X_n(s)-X_{n-1}(s)\|\mathrm{d}s$$

按此递推关系可得

$$\|X_{n+1}(t)-X_n(t)\|$$

$$\leqslant M^n\int_{t_0}^{t}\mathrm{d}s\int_{t_0}^{s}\mathrm{d}s_1\cdots\int_{t_0}^{s_{n-2}}\|X_1(s_{n-1})-X_0(s_{n-1})\|\mathrm{d}s_{n-1}$$

但

$$\|X_1(s_{n-1})-X_0(s_{n-1})\|\leqslant\int_{t_0}^{s_{n-1}}\|A(s_n)\|\|C\|\mathrm{d}s_n$$

$$\leqslant M\|C\|\int_{t_0}^{s_{n-1}}\mathrm{d}s_n=M\|C\|(s_{n-1}-t_0)$$

故

$$\|X_{n+1}(t)-X_n(t)\|\leqslant M^{n+1}\|C\|\frac{(t-t_0)^{n+1}}{(n+1)!},\quad(t_0\leqslant t\leqslant t_1)$$

因为

$$\sum_{n=0}^{\infty}\|X_{n+1}(t)-X_n(t)\|\leqslant\sum_{n=0}^{\infty}\frac{M^{n+1}(t-t_0)^{n+1}\|C\|}{(n+1)!}$$

$$=\mathrm{e}^{M(t-t_0)}\|C\|\leqslant\mathrm{e}^{M(t_1-t_0)}\|C\|$$

所以，函数矩阵级数

$$\sum_{n=0}^{\infty}(X_{n+1}(t)-X_n(t))$$

在 $t_0\leqslant t\leqslant t_1$ 上一致收敛。设此级数的和函数为 $X(t)$，即

$$\lim_{N\to\infty}\sum_{n=0}^{N}(X_{n+1}(t)-X_n(t))=X(t)$$

又

$$\sum_{n=0}^{N}(X_{n+1}(t)-X_n(t))=X_{N+1}(t)$$

故有

$$\lim_{N\to\infty}X_{N+1}(t)=X(t)$$

在式(10-2)中，命 $n\to\infty$，便得

$$X(t)=C+\int_{t_0}^{t}A(s)X(s)\mathrm{d}s \tag{10-3}$$

易知，式(10-3)为方程(10-1)满足初始条件 $X(t_0)=C$的解。

下面证明解的唯一性。

设

$$Y(t)=C+\int_{t_0}^{t}A(s)Y(s))\mathrm{d}s$$

为满足初始条件 $Y(t_0)=C$ 的方程(10-1)的另外一个解，则

$$X(t)-Y(t)=\int_{t_0}^{t}A(s)(X(s)-Y(s))\mathrm{d}s$$

$$\|X(t)-Y(t)\|\leqslant\int_{t_0}^{t}\|A(s)\|\|X(s)-Y(s)\|\mathrm{d}s$$

多次应用上式，有

$$\begin{aligned}\|X(t)-Y(t)\|&\leqslant\int_{t_0}^{t}\|A(s)\|\|X(s)-Y(s)\|\mathrm{d}s\\&\leqslant M\int_{t_0}^{t}\mathrm{d}s\int_{t_0}^{s}\|A(s_1)\|\|X(s_1)-Y(s_1)\|\mathrm{d}s_1\\&\leqslant M^2\int_{t_0}^{t}\mathrm{d}s\int_{t_0}^{s}\|X(s_1)-Y(s_1)\|\mathrm{d}s_1\\&\cdots\cdots\\&\leqslant M^{n+1}\int_{t_0}^{t}\mathrm{d}s\int_{t_0}^{s}\mathrm{d}s_1\cdots\int_{t_0}^{s_{n-1}}\|X(s_n)-Y(s_n)\|\mathrm{d}s_n\end{aligned}$$

设 $h=\max\limits_{t_0\leqslant t\leqslant t_1}\|X(t)-Y(t)\|$，代入上式，得

$$\begin{aligned}\|X(t)-Y(t)\|&\leqslant M^{n+1}\int_{t_0}^{t}\mathrm{d}s\int_{t_0}^{s}\mathrm{d}s_1\cdots\int_{t_0}^{s_{n-1}}h\mathrm{d}s_n\\&=M^{n+1}h\frac{(t-t_0)^{n+1}}{(n+1)!}\end{aligned}$$

而

$$\lim_{n\to\infty}\frac{M^{n+1}(t-t_0)^{n+1}}{(n+1)!}=0$$

故

$$\|X(t)-Y(t)\|\leqslant 0$$

即

$$X(t)\equiv Y(t)$$ ∎

把

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=A(t)X(t)$$

两端转置，得

$$\frac{\mathrm{d}X^{\mathrm{T}}(t)}{\mathrm{d}t}=X^{\mathrm{T}}(t)A^{\mathrm{T}}(t)$$

因此，形如

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=X(t)A(t),\quad X(t_0)=C$$

的矩阵微分方程的解也存在且唯一。

定理 10.1.2 当 $\det C\neq 0$时，方程(10-1)的任何一个解都是非奇异的，即 $\det X(t)\neq 0$，$t_0\leqslant t\leqslant t_1$。

〔证明〕 首先，我们证明当 C 为非奇异矩阵时，方程(10-1)的解满足等式

$$\det X(t)=\det C\cdot \mathrm{e}^{\int_{t_0}^{t}\mathrm{tr}A(t)\mathrm{d}t} \tag{10-4}$$

等式(10-4)称为 Jacobi 等式。

设 $X(t)=(x_{ij}(t))_{n\times n}$，则

$$\frac{\mathrm{d}|X(t)|}{\mathrm{d}t}=\sum_{i=1}^{n}\begin{vmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t)\\ \cdots\cdots & \cdots & \cdots & \cdots\\ \dfrac{\mathrm{d}x_{i1}(t)}{\mathrm{d}t} & \dfrac{\mathrm{d}x_{i2}(t)}{\mathrm{d}t} & \cdots & \dfrac{\mathrm{d}x_{in}(t)}{\mathrm{d}t}\\ \cdots\cdots & \cdots & \cdots & \cdots\\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t)\end{vmatrix}$$

于是，有

$$\begin{vmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t)\\ \cdots & \cdots & \cdots & \cdots\\ \dfrac{\mathrm{d}x_{i1}(t)}{\mathrm{d}t} & \dfrac{\mathrm{d}x_{i2}(t)}{\mathrm{d}t} & \cdots & \dfrac{\mathrm{d}x_{in}(t)}{\mathrm{d}t}\\ \cdots & \cdots & \cdots & \cdots\\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t)\end{vmatrix}=\begin{vmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t)\\ \cdots & \cdots & \cdots & \cdots\\ \sum a_{ij}x_{j1} & \sum a_{ij}x_{j2} & \cdots & \sum a_{ij}x_{jn}\\ \cdots & \cdots & \cdots & \cdots\\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t)\end{vmatrix}$$

$$=\begin{vmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t)\\ \cdots & \cdots & \cdots & \cdots\\ a_{i1}x_{11} & a_{i1}x_{12} & \cdots & a_{i1}x_{1n}\\ \cdots & \cdots & \cdots & \cdots\\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t)\end{vmatrix}+\begin{vmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t)\\ \cdots & \cdots & \cdots & \\ a_{i2}x_{21} & a_{i2}x_{22} & \cdots & a_{i2}x_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t)\end{vmatrix}$$

$$
+\cdots+\begin{vmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t) \\ \cdots & \cdots & \cdots & \cdots \\ a_{ii}x_{i1} & a_{ii}x_{i2} & \cdots & a_{ii}x_{in} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t) \end{vmatrix}+\cdots+\begin{vmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t) \\ \cdots & \cdots & \cdots & \cdots \\ a_{in}x_{n1} & a_{in}x_{n2} & \cdots & a_{in}x_{nn} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t) \end{vmatrix}
$$

$$
=a_{ii}\begin{vmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t) \\ \cdots & \cdots & \cdots & \cdots \\ x_{i1}(t) & x_{i2}(t) & \cdots & x_{in}(t) \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nn}(t) \end{vmatrix}=a_{ii}|X(t)|
$$

于是

$$
\frac{\mathrm{d}|X(t)|}{\mathrm{d}t}=\sum_{i=1}^{n}a_{ii}|X(t)|=|X(t)|\sum_{i=1}^{n}a_{ii}=|X(t)|\operatorname{tr}A(t)
$$

把上式改写成

$$
\frac{\mathrm{d}|X(t)|}{|X(t)|}=\operatorname{tr}A(t)\mathrm{d}t
$$

对上式两端从 0 到 t 积分

$$
\int_0^t\frac{\mathrm{d}|X(t)|}{|X(t)|}=\int_0^t\operatorname{tr}A(s)\mathrm{d}s
$$

有

$$
\ln|X(t)|\Big|_{t=0}^{t=t}=\int_0^t\operatorname{tr}A(s)\mathrm{d}s
$$

即

$$
\ln\frac{|X(t)|}{|C|}=\int_0^t\operatorname{tr}A(s)\mathrm{d}s
$$

于是

$$
|X(t)|=|C|\mathrm{e}^{\int_0^t\operatorname{tr}A(s)\mathrm{d}s}
$$

这就是我们要证明的 Jacobi 等式。

由 Jacobi 等式知，只要 $|C|\neq0$,就有 $|X(t)|\neq0$。

定理 10.1.3 设方程

$$
\begin{cases}\dfrac{\mathrm{d}X(t)}{\mathrm{d}t}=A(t)X(t)\\ X(t_0)=C_1\end{cases},\quad\begin{cases}\dfrac{\mathrm{d}X(t)}{\mathrm{d}t}=A(t)X(t)\\ X(t_0)=C_2\end{cases}
$$

的解分别为 $X_1(t)$和 $X_2(t)$，则

$$X_2(t)=X_1(t)T,\quad T=C_1^{-1}C_2$$

〔证明〕 命 $X_3(t)=X_1(t)C_1^{-1}C_2=X_1(t)T$，则

$$\frac{\mathrm{d}X_3(t)}{\mathrm{d}t}=\frac{\mathrm{d}X_1(t)}{\mathrm{d}t}T=A(t)X_1(t)T=A(t)X_3(t)$$

且

$$X_3(t_0)=X_1(t_0)T=C_1T=C_2$$

根据解的唯一性，有 $X_3(t)=X_2(t)$，于是

$$X_2(t)=X_1(t)T$$

▮

称方程

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=A(t)X(t),\quad X(t_0)=I$$

的解 $X=X_0(t)$为 $A(t)$的基本解矩阵。由定理 10.1.3 可知，满足其他初始条件的解矩阵，都可以由基本解矩阵表示。

例 10.1.1 设 $A(t)$是连续矩阵函数，且对任何 s，t 在 $A(t)$所定义的区间内有

$$A(s)A(t)=A(t)A(s)$$

试证明

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=A(t)X(t),\quad X(0)=I$$

的解为

$$X(t)=\exp\left(\int_0^t A(s)\mathrm{d}s\right) \tag{10 5}$$

〔证明〕 令

$$B(t)=\int_0^t A(s)\mathrm{d}s$$

则有

$$\frac{\mathrm{d}B(t)}{\mathrm{d}t}=A(t) \tag{10-6}$$

又

$$B(t)\frac{\mathrm{d}B(t)}{\mathrm{d}t}=\frac{\mathrm{d}B(t)}{\mathrm{d}t}B(t) \tag{10-7}$$

这是因为

$$B(t)\frac{\mathrm{d}B(t)}{\mathrm{d}t}=\int_0^t A(s)\mathrm{d}s\cdot A(t)=\int_0^t A(s)A(t)\mathrm{d}s$$

$$=\int_0^t A(t)A(s)\mathrm{d}s=A(t)\int_0^t A(s)\mathrm{d}s=\frac{\mathrm{d}B(t)}{\mathrm{d}t}B(t)$$

于是

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\exp\int_0^t A(s)\mathrm{d}s\right)=\frac{\mathrm{d}}{\mathrm{d}t}(\exp B(t))$$

$$=\frac{\mathrm{d}}{\mathrm{d}t}\left(I+B(t)+\frac{1}{2!}B^2(t)+\cdots\right)$$

$$=\frac{\mathrm{d}B(t)}{\mathrm{d}t}+\frac{1}{2!}\left(\frac{\mathrm{d}B(t)}{\mathrm{d}t}B(t)+B(t)\frac{\mathrm{d}B(t)}{\mathrm{d}t}\right)+\cdots$$

由式(10-6)和式(10-7)，可知

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=A(t)\exp(B(t))=A(t)X(t)$$

且

$$X(0)=\exp\left(\int_0 A(s)\mathrm{d}s\right)=I$$

例 10.1.2 求

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=\begin{pmatrix} t & 1 \\ 2 & t \end{pmatrix}X(t),\quad X(0)=I$$

的解

〔解〕 因为

$$A(t)=\begin{pmatrix} t & 1 \\ 2 & t \end{pmatrix}=t\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}+\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}=tI+B$$

所以

$$A(t)A(s)=A(s)A(t)$$

由例 10.1.1，知

$$X(t)=\exp\left(\int_0^t A(s)\mathrm{d}s\right)$$

$$=\exp\left(\int_0^t (sI+B)\mathrm{d}s\right)$$

$$=\exp\left(\frac{1}{2}t^2I+Bt\right)$$

于是

$$X(t)=\begin{pmatrix} e^{-\frac{1}{2}t^2}\operatorname{ch}\sqrt{2}\,t & \frac{1}{\sqrt{2}}e^{-\frac{1}{2}t^2}\operatorname{sh}\sqrt{2}\,t \\ \sqrt{2}\,e^{-\frac{1}{2}t^2}\operatorname{sh}\sqrt{2}\,t & e^{-\frac{1}{2}t^2}\operatorname{ch}\sqrt{2}\,t \end{pmatrix}$$

定理 10.1.4 若 $A(t)=A$ 是常数矩阵，那末方程(10-1)的解为

$$X(t)=e^{A(t-t_0)}C \tag{10-8}$$

事实上，把式(10-8)代入方程(10-1)验算，即知定理为真。 ▮

推论 10.1.1 若 A 为常数矩阵，则方程

$$\frac{dX(t)}{dt}=X(t)A, \quad X(t_0)=C$$

的解为

$$X(t)=Ce^{A(t-t_0)}$$

由推论 10.1.1 知，当 $X(0)=I$ 时，方程

$$\frac{dX(t)}{dt}=X(t)A \quad (A\text{ 为常数矩阵})$$

的解为 $X(t)=e^{At}$。而指数函数 e^{At} 满足 $e^{A(t+s)}=e^{At}e^{As}$，也就是说，$X(t)=e^{At}$ 满足矩阵方程

$$X(t+s)=X(t)X(s),\ -\infty<t、s<+\infty \tag{10-9}$$

反之，可以证明，满足方程(10-9)的矩阵函数只能是矩阵指数函数。这就是下面的定理。

定理 10.1.5 设函数矩阵 $X(t)$ 对于任何有限的 t，可以求导且满足方程

$$X(t+s)=X(t)X(s),\ -\infty<t、s<+\infty$$

$$X(0)=I$$

则

$$X(t)=e^{At}$$

〔证明〕 对于任何有限的 t，有

$$I=X(t-t)=X(t)X(-t)$$

这表明 $X(t)$ 有逆矩阵，且 $X^{-1}(t)=X(-t)$。

又由 $X(t+s)=X(t)X(s)$，有

$$\frac{dX(s+t)}{d(s+t)}=\frac{dX(t+s)}{d(t+s)}$$

而

$$\frac{dX(s+t)}{d(s+t)}=\frac{dX(s+t)}{d(s+t)}\ \frac{d(s+t)}{dt}$$

$$= \frac{dX(s+t)}{dt} = X(s)\frac{dX(t)}{dt}$$

同样可得

$$\frac{dX(t+s)}{d(t+s)} = X(t)\frac{dX(s)}{ds}$$

于是

$$X(s)\frac{dX(t)}{dt} = X(t)\frac{dX(s)}{ds} \qquad (10\text{-}10)$$

在式(10-10)两端左乘 $X^{-1}(t)X^{-1}(s)$，有

$$X^{-1}(t)\frac{dX(t)}{dt} = X^{-1}(t)X^{-1}(s)X(t)\frac{dX(s)}{ds}$$

$$= X(-t-s+t)\frac{dX(s)}{ds}$$

$$= X(-s)\frac{dX(s)}{ds} = X^{-1}(s)\frac{dX(s)}{ds}$$

即

$$X^{-1}(t)\frac{dX(t)}{dt} = X^{-1}(s)\frac{dX(s)}{ds} \qquad (10\text{-}11)$$

式(10-11)表明 $X^{-1}(t)\frac{dX(t)}{dt}$ 是一个常数矩阵，设为 A，即

$$X^{-1}(t)\frac{dX(t)}{dt} = A$$

于是

$$\frac{dX(t)}{dt} = X(t)A, \quad X(0) = I \qquad (10\text{-}12)$$

这表明 $X(t)$是方程(10-12)的解，且

$$X(t) = e^{At}$$

§10.2 线性齐次向量微分方程

形如

$$\begin{cases} \frac{dx_1}{dt} = a_{11}(t)x_1(t) + a_{12}(t)x_2(t) + \cdots + a_{1n}(t)x_n(t) \\ \frac{dx_2}{dt} = a_{21}(t)x_1(t) + a_{22}(t)x_2(t) + \cdots + a_{2n}(t)x_n(t) \end{cases}$$

$$\begin{cases} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{dx_n}{dt}=a_{n1}(t)x_1(t)+a_{n2}(t)x_2(t)+\cdots+a_{nn}(t)x_n(t) \end{cases}$$

的微分方程组与初始条件

$$x_1(t_0)=x_{10},\ x_2(t_0)=x_{20},\ \cdots,\ x_n(t_0)=x_{n0}$$

可以写成下面的形式

$$\frac{d\boldsymbol{x}(t)}{dt}=A(t)\boldsymbol{x}(t),\ \boldsymbol{x}(t_0)=\boldsymbol{x}_0 \tag{10-13}$$

其中 $\boldsymbol{x}(t)=(x_1(t),\ x_2(t),\ \cdots,\ x_n(t))^T$,

$$A(t)=\begin{pmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{pmatrix}$$

在 $A(t)$分段连续的条件下，方程(10-13)的解存在且唯一，同时解对 t_0、$\boldsymbol{x}_0$ 是连续的。用

$$\boldsymbol{x}=\varphi(t;\ \boldsymbol{x}_0,\ t_0) \tag{10-14}$$

表示方程(10-13)满足初始条件 $\boldsymbol{x}(t_0)=\boldsymbol{x}_0$ 的解。

下面我们来讨论解 $\varphi(t;\ \boldsymbol{x}_0,\ t_0)$的性质。

命题 10.2.1 $\varphi(t_0;\ \boldsymbol{x}_0,\ t_0)=0$ 的充要条件是 $\varphi(t;\ \boldsymbol{0},\ t_0)=0$ 。

〔证明〕 $\varphi(t_0;\ \boldsymbol{x}_0,\ t_0)=0$ 表明方程(10-13)的解是零向量，即

$$\boldsymbol{x}=\boldsymbol{x}(t)=\varphi(t;\ \boldsymbol{x}_0,\ t_0)=0$$

故有

$$\boldsymbol{x}_0=\boldsymbol{x}(t_0)=0$$

反之，若 $\boldsymbol{x}_0=0$，则 $\boldsymbol{x}=0$ 满足

$$\frac{d\boldsymbol{x}}{dt}=A\boldsymbol{x},\ \boldsymbol{x}(t_0)=0$$

即

$$\varphi(t_0;\ \boldsymbol{x}_0,\ t_0)=0$$ 】

命题 10.2.2 $\varphi(t;\ \boldsymbol{x}_0,\ t_0)$对初始条件是线性的，即

$$\varphi(t;\ a\boldsymbol{x}_1+b\boldsymbol{x}_2,\ t_0)=a\varphi(t;\ \boldsymbol{x}_1,\ t_0)+b\varphi(t;\ \boldsymbol{x}_2,\ t_0)$$

其中 a、b 为两个任意常数。

〔证明〕 $\varphi(t;\ \boldsymbol{x}_1,\ t_0)$、$\varphi(t;\ \boldsymbol{x}_2,\ t_0)$分别是方程

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=A(t)\boldsymbol{x}(t),\ \boldsymbol{x}(t_0)=\boldsymbol{x}_1$$

和

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=A(t)\boldsymbol{x}(t),\ \boldsymbol{x}(t_0)=\boldsymbol{x}_2$$

的解。而方程

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=A(t)\boldsymbol{x}(t),\ \boldsymbol{x}(t_0)=a\boldsymbol{x}_1+b\boldsymbol{x}_2 \tag{10-15}$$

的解为

$$\boldsymbol{x}=\varphi(t;\ a\boldsymbol{x}_1+b\boldsymbol{x}_2,\ t_0)$$

下面我们来验证

$$\boldsymbol{x}=a\varphi(t;\ \boldsymbol{x}_1,\ t_0)+b\varphi(t;\ \boldsymbol{x}_2,\ t_0)$$

也是方程(10-15)的解。

事实上，

$$\frac{\mathrm{d}}{\mathrm{d}t}[a\varphi(t;\ \boldsymbol{x}_1,\ t_0)+b\varphi(t;\ \boldsymbol{x}_2,\ t_0)]$$

$$=a\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t;\ \boldsymbol{x}_1,\ t_0)+b\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t;\ \boldsymbol{x}_2,\ t_1)$$

$$=aA\varphi(t;\ \boldsymbol{x}_1,\ t_0)+bA\varphi(t;\ \boldsymbol{x}_2,\ t_0)$$

$$=A(a\varphi(t;\ \boldsymbol{x}_1,\ t_0)+b\varphi(t;\ \boldsymbol{x}_2,\ t_0))$$

且

$$a\varphi(t_0;\ \boldsymbol{x}_1,\ t_0)+b\varphi(t_0;\ \boldsymbol{x}_2,\ t_0)=a\boldsymbol{x}_1+b\boldsymbol{x}_2$$

由解的唯一性，有

$$\varphi(t;\ a\boldsymbol{x}_1+b\boldsymbol{x}_2,\ t_0)=a\varphi(t;\ \boldsymbol{x}_1,\ t_0)+b\varphi(t;\ \boldsymbol{x}_2,\ t_0)$$ ∎

命题 10.2.3 设 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 为 n 维常数向量，则

$$\varphi(t;\ \boldsymbol{\alpha}_1,\ t_0),\ \varphi(t;\ \boldsymbol{\alpha}_2,\ t_0),\ \cdots,\ \varphi(t;\ \boldsymbol{\alpha}_n,\ t_0)$$

线性无关的充要条件是 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 线性无关。

〔证明〕 由命题 10.2.2，对常数 k_1，k_2，…，k_n 有

$$\sum_{i=1}^{n}k_i\varphi(t;\ \boldsymbol{\alpha}_i,\ t_0)=\varphi\left(t;\ \sum_{i=1}^{n}k_i\boldsymbol{\alpha}_i,\ t_0\right)$$

如果上式等于零，则由命题 10.2.1，知

$$\sum_{i=1}^{n}k_i\boldsymbol{\alpha}_i=0$$

于是 $\varphi(t;\ \boldsymbol{a}_1,\ t_0)$, $\varphi(t;\ \boldsymbol{a}_2,\ t_0)$, …, $\varphi(t;\ \boldsymbol{a}_n,\ t_0)$线性无关的充要条件是$\boldsymbol{a}_1, \boldsymbol{a}_2$, …, $\boldsymbol{a}_n$线性无关。]

命题 10.2.4 设 $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, …, $\boldsymbol{e}_n$是R^n中的基本向量组，即$\boldsymbol{e}_i=(\underbrace{0,0,\cdots,0}_{(i-1)\text{个}},\ 1,\ 0,\ \cdots,\ 0)^{\mathrm{T}}$，则

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=A(t)\boldsymbol{x}(t),\ \boldsymbol{x}(t_0)=\boldsymbol{x}_0 \tag{10-16}$$

的任何一个解 $\boldsymbol{x}=\boldsymbol{x}(t)$可以表示成

$$\boldsymbol{x}(t)=\sum_{i=1}^{n}k_i\varphi(t;\ \boldsymbol{e}_i,\ t_0) \tag{10-17}$$

〔证明〕 设向量 $\boldsymbol{x}_0$ 为

$$\boldsymbol{x}_0=\sum_{i=1}^{n}k_i\boldsymbol{e}_i$$

于是

$$\boldsymbol{x}(t)=\varphi(t;\ \boldsymbol{x}_0,\ t_0)=\sum_{i=1}^{n}k_i\varphi(t;\ \boldsymbol{e}_i,\ t_0)$$]

由命题10.2.3和命题10.2.4知，$\varphi(t;\ \boldsymbol{e}_1,\ t_0)$, $\varphi(t;\ \boldsymbol{e}_2,\ t_0)$,…, $\varphi(t;\ \boldsymbol{e}_n,t_0)$线性无关。方程(10-16)的任何一个解都可以由这 n 个列向量线性表出。

构造一个 n 阶矩阵

$$\boldsymbol{\Phi}(t,\ t_0)\triangleq(\varphi(t;\ \boldsymbol{e}_1,\ t_0),\ \varphi(t;\ \boldsymbol{e}_2,\ t_0),\ \cdots,\ \varphi(t;\ \boldsymbol{e}_n,\ t_0)) \tag{10-18}$$

它的第 i 列由列向量 $\varphi(t;\ \boldsymbol{e}_i,\ t_0)$ 构成。由命题 10.2.4，方程(10-16)的解可以写成

$$\boldsymbol{x}(t)=\boldsymbol{\Phi}(t,\ t_0)\boldsymbol{x}(t_0) \tag{10-19}$$

上式表明，经过矩阵 $\boldsymbol{\Phi}(t,\ t_0)$作用后，状态从 $\boldsymbol{x}(t_0)$转移到 $\boldsymbol{x}(t)$。在工程问题中，通常把 $\boldsymbol{\Phi}(t,\ t_0)$称为 $A(t)$的**状态转移矩阵**。

在前面的讨论中，把 t_0 看成是某个固定的常数，实际上可以把它看作是在$(-\infty,\ \infty)$范围内变化的量，并把它记成 s。因此，状态转移矩阵是两个变量 t，s 的函数。

定理 10.2.1 状态转移矩阵 $\boldsymbol{\Phi}(t,\ s)$是矩阵微分方程

$$\frac{\partial\boldsymbol{\Phi}(t,\ t_0)}{\partial t}=A(t)\boldsymbol{\Phi}(t,\ s),\ \boldsymbol{\Phi}(s,\ s)=I \tag{10-20}$$

的唯一解。

〔证明〕 因为

$$\frac{\partial\boldsymbol{\Phi}(t,\ s)}{\partial t}\left(\frac{\partial\varphi(t;\ \boldsymbol{e}_1,\ s)}{\partial t},\ \frac{\partial\varphi(t;\ \boldsymbol{e}_2,\ s)}{\partial t},\ \cdots,\ \frac{\partial\varphi(t;\ \boldsymbol{e}_n,\ s)}{\partial t}\right)$$

$$=(A\varphi(t;\ \boldsymbol{e}_1,\ s),\ A\varphi(t;\ \boldsymbol{e}_2,\ s),\ \cdots,\ A\varphi(t;\ \boldsymbol{e}_n,\ s))$$

$$=A(\varphi(t;\ \boldsymbol{e}_1,\ s),\varphi(t;\ \boldsymbol{e}_2,\ s),\ \cdots,\ \varphi(t;\ \boldsymbol{e}_n,\ s))$$

$$=A\Phi(t, s)$$

且

$$\Phi(s, s)=(\varphi(s; e_1, s), \varphi(s; e_2, s), \cdots, \varphi(s; e_n, s))$$
$$=(e_1, e_2, \cdots, e_n)=I$$

所以，状态转移矩阵 $\Phi(t, s)$是方程(10-20)的解。由定理 10.1.1 知解是唯一的。 ∎

推论 10.2.1 当 A 为常数矩阵时，方程

$$\frac{dx}{dt}=Ax(t), \quad x(t_0)=x_0$$

的解有表达式

$$x=e^{A(t-t_0)}x_0$$

§10.3 状态转移矩阵

由§10.2知，矩阵微分方程

$$\frac{\partial\Phi(t, s)}{\partial t}=A(t)\Phi(t, s), \quad \Phi(s, s)=I \tag{10-21}$$

的解，称为$A(t)$的**状态转移矩阵**。这一节我们来讨论状态转移矩阵的性质。

定理 10.3.1 $\Phi(t, t_1)\Phi(t_1, s)=\Phi(t, s)$

〔证明〕 由式(10-19)，有

$$(\Phi(t, t_1)\Phi(t_1, s))x(s)=\Phi(t, t_1)\Phi(t_1, s)x(s)$$
$$=\Phi(t, t_1)x(t_1)=x(t)$$

又

$$x(t)=\Phi(t, s)x(s)$$

根据状态转移矩阵的唯一性，得

$$\Phi(t, t_1)\Phi(t_1, s)=\Phi(t, s)$$ ∎

定理 10.3.2 $(\Phi(t, s))^{-1}=\Phi(s, t)$

〔证明〕 因为

$$\Phi(t, s)\Phi(s, t)=\Phi(t, t)=I$$

故有

$$(\Phi(t, s))^{-1}=\Phi(s, t)$$ ∎

因为 $\Phi(t, s)$是矩阵微分方程(10-21)的解，所以下面的定理是显然的。

定理 10.3.3 $\Phi^T(t, s)$满足矩阵微分方程

$$\frac{\partial \Phi^T(t, s)}{\partial t} = \Phi^T(t, s)A^T(t)$$

定理 10.3.4 $\Phi(t, s)$满足矩阵微分方程

$$\frac{\partial \Phi(t, s)}{\partial t} = -\Phi(t, s)A(s), \quad \Phi(s, s) = I \qquad (10\text{-}22)$$

〔证明 由定理 10.3.2 和函数矩阵 $A(x)$与 $A^{-1}(x)$间的导数关系*，有

$$\frac{\partial \Phi(t, s)}{\partial s} = -\frac{\partial [\Phi(s, t)]^{-1}}{\partial s}$$

$$= -[\Phi(s, t)]^{-1}\frac{\partial \Phi(s, t)}{\partial s}[\Phi(s, t)]^{-1}$$

$$= -[\Phi(s, t)]^{-1}A(s)\Phi(s, t)][\Phi(s, t)]^{-1}$$

$$= -[\Phi(s, t)]^{-1}A(s)$$ 〕

定理 10.3.5 设$A(t)$的状态转移矩阵为$\Phi(t, s)$，$-A^T(t)$的状态转移矩阵为 $\Psi(t,s)$，则

$$\Psi(t, s) = [\Phi^T(t, s)]^{-1}$$

〔证明〕 把式(10-22)两端转置，得

$$\frac{\partial \Phi^T(t, s)}{\partial s} = -A^T(s)\Phi^T(t, s)$$

又

$$[\Phi^T(s, t)]^{-1} = [\Phi^{-1}(s, t)]^T = \Phi^T(t, s)$$

于是

$$\frac{\partial [\Phi^T(s, t)]^{-1}}{\partial s} = \frac{\partial \Phi^T(t, s)}{\partial s} = -A^T(s)\Phi^T(t, s)$$

$$= -A^T(s)[\Phi^T(s, t)]^{-1}$$

上式表明，$[\Phi^T(s, t)]^{-1}$ 是 $-A^T(s)$的状态转移矩阵，所以

$$\Psi(t, s) = [\Phi^T(s, t)]^{-1}$$ 〕

§10.4 线性非齐次向量微分方程

定理 10.4.1 线性非齐次向量微分方程

---

* 如果函数矩阵$A(x)$与 $A^{-1}(x)$都有导数，则

$$\frac{dA^{-1}(x)}{dx} = -A^{-1}(x)\frac{dA(x)}{dx}A^{-1}(x)$$

$$\frac{dx(t)}{dt}=A(t)x(t)+f(t),\quad x(t_0)=c \tag{10-23}$$

的唯一解可表示成

$$x(t)=\Phi(t,\ t_0)\left[c+\int_{t_0}^{t}\Phi(t_0,\ \tau)f(\tau)d\tau\right] \tag{10-24}$$

其中$A(t)$是n阶分段连续的函数矩阵，$f(t)$是n维函数列向量，$x(t)$是n维未知列向量，$\Phi(t,\ t_0)$是$A(t)$的状态转移矩阵。

〔证明〕 把式(10-24)代入式(10-23)，有

$$\begin{aligned}\frac{dx(t)}{dt}&=\frac{\partial\Phi(t,\ t_0)}{\partial t}\left[c+\int_{t_0}^{t}\Phi(t_0,\ \tau)f(\tau)d\tau\right]\\&\quad+\Phi(t,\ t_0)\Phi(t_0,\ t)f(t)\\&=A(t)\Phi(t,\ t_0)\left[c+\int_{t_0}^{t}\Phi(t_0,\ \tau)f(\tau)d\tau\right]+f(t)\\&=A(t)x(t)+f(t)\end{aligned}$$

且

$$x(t_0)=\Phi(t,\ t_0)\left[c+\int_{t_0}^{t_0}\Phi(t_0,\ \tau)f(\tau)d\tau\right]=c$$ ▮

推论 10.4.1 若$A(t)$为常数矩阵，则方程(10-23)的唯一解可以表示为

$$x(t)=e^{A(t-t_0)}c+e^{A(t-t_0)}\int_{t_0}^{t}e^{A(t-\tau)}f(\tau))d\tau \tag{10-25}$$

§10.5 形如 dX/dt=A(t)X(t)+X(t)B(t)的方程

形如

$$\frac{dX(t)}{dt}=A(t)X(t)+X(t)B(t),\quad X(t_0)=C \tag{10-26}$$

的矩阵微分方程解的存在和唯一性定理的证明与定理 10.1.1 的类似，我们就不再赘述了。

定理 10.5.1 对任意的$t\in R$，方程(10-26)的唯一解可以表示为

$$X(t,\ C,\ t_0)=\Phi_1(t,\ t_0)C\Phi_2(t,\ t_0) \tag{10-27}$$

其中$\Phi_1(t,\ t_0)$，$\Phi_2(t,\ t_0)$分别是由

$$\frac{\partial\Phi_1(t,\ t_0)}{\partial t}=A(t)\Phi_1(t,t_0),\quad \Phi_1(t_0,\ t_0)=I \tag{10-28}$$

和

$$\frac{\partial\Phi_2(t,\ t_0)}{\partial t}=\Phi_2(t,\ t_0)B(t),\quad \Phi_2(t_0,\ t_0)=I \tag{10-29}$$

所定义的状态转移矩阵。

将式(10-27)对 t 求偏导数，然后应用式(10-28)和式(10-29)即可证明定理 10.5.1。

推论 10.5.1 若 $A(t)$和 $B(t)$是常数矩阵，则方程(10-26)的解可以表示为

$$X(t,\ t_0,\ C)=\mathrm{e}^{A(t-t_0)}C\mathrm{e}^{B(t-t_0)} \tag{10-30}$$

类似地，我们可以得到形如

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=A(t)X(t)+X(t)B(t)+F(t),\qquad X(t_0)=C \tag{10-31}$$

的矩阵微分方程解的表达式。

定理 10.5.2 方程(10-31)的解可以表示为

$$X(t,\ t_0,\ C)=\Phi_1(t,\ t_0)C\,\Phi_2(t,\ t_0)+\int_{t_0}^{t}\Phi_1(t,\tau)F(\tau)\Phi_2(t,\tau)\mathrm{d}\tau$$

其中 $\Phi_1(t,\ t_0)$、$\Phi_2(t,\ t_0)$分别为式(10-28)和式(10-29)定义的状态转移矩阵，$A(t)$、$B(t)$、$F(t)$均为分段连续的 n 阶函数矩阵。

推论 10.5.2 如果 $B(t)=D$，$A(t)=D^T$ 为常数矩阵，$F(t)=Q$ 为常数对称矩阵，则方程

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=D^TX(t)+X(t)D+Q,\quad X(t_0)=C$$

的唯一解可表示为

$$X(t,\ t_0,\ C)=\mathrm{e}^{D^T(t-t_0)}C\mathrm{e}^{D(t-t_0)}+\int_{t_0}^{t}\mathrm{e}^{D^T(t-\tau)}Q\mathrm{e}^{D(t-\tau)}\mathrm{d}\tau$$

习　题

10-1 试求下列矩阵微分方程的解

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t}=AX(t),\qquad X(0)=I$$

(1) $A=\begin{pmatrix}0 & 1 & t\\ 0 & 0 & 1\\ 0 & 0 & 0\end{pmatrix}$　　(2) $A=\begin{pmatrix}1 & -t\\ t & \text{[illegible]}\end{pmatrix}$

(3) $A=\begin{pmatrix}3 & -2\\ 4 & -3\end{pmatrix}$

10-3 求未知向量 $x(t)$，它满足

(1) $\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\begin{pmatrix}0 & 1 & t\\ 0 & 0 & 1\\ 0 & 0 & 0\end{pmatrix}\boldsymbol{x}(t)+\begin{pmatrix}1\\ t\\ t^2\end{pmatrix},\qquad \boldsymbol{x}(0)=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$

(2) $\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\begin{pmatrix}1 & -t\\ t & 1\end{pmatrix}\boldsymbol{x}(t)+\mathrm{e}^t\begin{pmatrix}1\\ -1\end{pmatrix},\qquad \boldsymbol{x}(0)=\begin{pmatrix}0\\ 0\end{pmatrix}$

(2) $\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\begin{pmatrix}3 & -2\\ 4 & -3\end{pmatrix}\boldsymbol{x}(t)+t\begin{pmatrix}1\\ 1\end{pmatrix},\qquad \boldsymbol{x}(0)=\begin{pmatrix}0\\ 0\end{pmatrix}$

10-3 证明推论 10.1.1。

10-4 证明推论 10.2.1。

10-5 证明定理 10.1.4。

10-6 证明推论 10.4.1。

10-7 证明定理 10.5.1。

10-8 证明推论 10.5.1。

10-9 证明定理 10.5.2。

10-10 证明推论 10.5.2。

第十一章 Kronecker积与矩阵代数方程

这一章我们讨论含有未知矩阵的矩阵代数方程。在讨论矩阵代数方程之前，我们先引入矩阵的Kronecker积的概念，并讨论它的一些基本性质。Kronecker积不仅在矩阵代数方程的研究中起着重要的作用，而且在其他方面也有许多应用。

本章讨论的内容有：Kronecker积，矩阵的Kronecker积与Kronecker和的特征值，矩阵的列展开与行展开，线性矩阵代数方程。

§11.1 Kronecker积

我们知道，在矩阵代数中曾定义过两个矩阵A和B的乘积AB，它要求A的列数必须等于B的行数，否则AB是没有意义的。下面我们来引入一种新的矩阵乘法运算，它对矩阵的行数和列数没有任何要求。

定义 11.1.1 设$A=(a_{ij})_{m\times n}$，$B=(b_{ij})_{p\times q}$，则称由

$$\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

所确定的$mp\times nq$矩阵为A与B的**Kronecker积**，或称A与B的直积，记作$A\otimes B$，即

$$A\otimes B=\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

例 11.1.1 设

$$X=(x_1, x_2, x_3)^T, \quad Y=(y_1, y_2)^T$$

则

$$X\otimes Y=\begin{pmatrix} x_1Y \\ x_2Y \\ x_3Y \end{pmatrix}=(x_1y_1, x_1y_2, x_2y_1, x_2y_2, x_3y_1, x_3y_2)^T$$

$$Y\otimes X=\begin{pmatrix} y_1X \\ y_2X \end{pmatrix}=(y_1x_1, y_1x_2, y_1x_3, y_2x_1, y_2x_2, y_2x_3)^T$$

由例 11.1.1 可见，矩阵的 Kronecker 积不满足交换律。

对单位矩阵，有

$$I_n \otimes I_m = I_m \otimes I_n = I_{mn}$$

不难验证，矩阵的 Kronecker 积满足下面的运算律：

（1） $k(A\otimes B)=kA\otimes B=A\otimes kB$，$k$ 为任意复数。

（2） 分配律：$(A+B)\otimes C=A\otimes C+B\otimes C$。

（3） 结合律：$(A\otimes B)\otimes C=A\otimes(B\otimes C)$。

下面我们来证明 Kronecker 积的一个重要性质，它在 Kronecker 积的研究中起着重要的作用。

定理 11.1.1 设 $A=(a_{ij})_{m\times n}$，$B=(b_{ij})_{s\times t}$，$C=(c_{ij})_{n\times p}$，$D=(d_{ij})_{t\times q}$，则

$$(A\otimes B)(C\otimes D)=AC\otimes BD \tag{11-1}$$

〔证明〕

$$(A\otimes B)(C\otimes D)=\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}\begin{pmatrix} c_{11}D & c_{12}D & \cdots & c_{1p}D \\ c_{21}D & c_{22}D & \cdots & c_{2p}D \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1}D & c_{n2}D & \cdots & c_{np}D \end{pmatrix}$$

$$=\begin{pmatrix} \left(\sum_{k=1}^{n}a_{1k}c_{k1}\right)BD & \left(\sum_{k=1}^{n}a_{1k}c_{k2}\right)BD & \cdots & \left(\sum_{k=1}^{n}a_{1k}c_{kp}\right)BD \\ \left(\sum_{k=1}^{n}a_{2k}c_{k1}\right)BD & \left(\sum_{k=1}^{n}a_{2k}c_{k2}\right)BD & \cdots & \left(\sum_{k=1}^{n}a_{2k}c_{kp}\right)BD \\ \cdots & \cdots & \cdots & \cdots \\ \left(\sum_{k=1}^{n}a_{mk}c_{k1}\right)BD & \left(\sum_{k=1}^{n}a_{mk}c_{k2}\right)BD & \cdots & \left(\sum_{k=1}^{n}a_{mk}c_{kp}\right)BD \end{pmatrix}$$

$$=(AC)\otimes(BD)$$

▌

推论 11.1.1 若 $A=(a_{ij})_{m\times m}$，$B=(b_{ij})_{n\times n}$，则

$$A\otimes B=(A\otimes I_n)(I_m\otimes B)=(I_m\otimes B)(A\otimes I_n)$$

定理 11.1.2 设 $A=(a_{ij})_{m\times n}$，$B=(b_{ij})_{p\times q}$，则

$$(A\otimes B)^T=A^T\otimes B^T$$

$$(A\otimes B)^H=A^H\otimes B^H$$

定理的证明留给读者。

定理 11.1.3 设 A、B 分别为 m 阶和 n 阶可逆矩阵，则 $A\otimes B$ 也为可逆矩阵，且

$$(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$$

〔证明〕 由定理 11.1.1，有

$$\begin{aligned}(A\otimes B)(A^{-1}\otimes B^{-1})&=AA^{-1}\otimes BB^{-1}\\&=I_m\otimes I_n=I_{mn}\end{aligned}$$

故 $$(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$$ ▌

定理 11.1.4 设 $A=(a_{ij})_{m\times m}$，$B=(b_{ij})_{n\times n}$，则

$$\operatorname{tr}(A\otimes B)=\operatorname{tr}A\cdot\operatorname{tr}B$$

〔证明〕 因为

$$A\otimes B=\begin{pmatrix}a_{11}B & a_{12}B & \cdots & a_{1m}B\\ a_{21}B & a_{22}B & \cdots & a_{2m}B\\ \cdots & \cdots & \cdots & \cdots\\ a_{m1}B & a_{m2}B & \cdots & a_{mm}B\end{pmatrix}$$

所以

$$\begin{aligned}\operatorname{tr}(A\otimes B)&=a_{11}\operatorname{tr}B+a_{22}\operatorname{tr}B+\cdots+a_{mn}\operatorname{tr}B\\&=(a_{11}+a_{22}+\cdots+a_{mm})\operatorname{tr}B\\&=\operatorname{tr}A\operatorname{tr}B\end{aligned}$$ ▌

定理 11.1.5 设 $A=(a_{ij})_{m\times n}$，$B=(b_{ij})_{p\times q}$，则

$$\operatorname{rank}(A\otimes B)=\operatorname{rank}(A)\operatorname{rank}(B)$$

〔证明〕 设矩阵 A 与 B 的标准形分别为 A_1 与 B_1，即

$$MAN=A_1,\quad PBQ=B_1 \tag{11-2}$$

其中 M、N、P、Q 分别为 m 阶、n 阶、p 阶和 q 阶非奇异矩阵，且

$$A_1=\begin{pmatrix}1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 1\end{pmatrix},\quad B_1=\begin{pmatrix}1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0\end{pmatrix}$$

A_1 中数 1 的个数为 $\operatorname{rank}(A)$，B_1 中数 1 的个数为 $\operatorname{rank}(B)$。

由式（11-2），有

$$A=M^{-1}A_1N^{-1},\quad B=P^{-1}B_1Q^{-1}$$

于是，由定理 11.1.1，有

$$\begin{aligned}A\otimes B&=(M^{-1}A_1N^{-1})\otimes(P^{-1}B_1Q^{-1})\\&=(M^{-1}\otimes P^{-1})(A_1\otimes B_1)(N^{-1}\otimes Q^{-1})\end{aligned}$$

由定理 11.1.3 知，$M^{-1}\otimes P^{-1}$、$N^{-1}\otimes Q^{-1}$ 均为非奇异矩阵，故

$$\text{rank}(A\otimes B)=\text{rank}(A_1\otimes B_1)$$

而 $A_1\otimes B_1$ 的秩为 $\text{rank}(A)\text{rank}(B)$，于是

$$\text{rank}(A\otimes B)=\text{rank}(A)\text{rank}(B)$$]

定理 11.1.6 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 是 n 个线性无关的 m 维列向量，$\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_q$ 是 q 个线性无关的 p 维列向量，则 nq 个 mp 维列向量

$$\boldsymbol{x}_i\otimes \boldsymbol{y}_j \quad (i=1, 2, \cdots, n;\ j=1, 2, \cdots, q) \tag{11-3}$$

线性无关。反之，若向量组（11-3）线性无关，则 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 和 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_q$ 均线性无关。

〔证明〕 设

$$\boldsymbol{x}_j=(a_{1j}, a_{2j}, \cdots, a_{mj})^T$$

$$\boldsymbol{y}_j=(b_{1j}, b_{2j}, \cdots, b_{pj})^T$$

令

$$A=[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]=(a_{ij})_{m\times n}$$

$$B=[\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_q]=(b_{ij})_{p\times q}$$

显然

$$\text{rank}(A)=n, \quad \text{rank}(B)=q$$

因为

$$A\otimes B=[\boldsymbol{x}_1\otimes\boldsymbol{y}_1, \boldsymbol{x}_1\otimes\boldsymbol{y}_2, \cdots, \boldsymbol{x}_1\otimes\boldsymbol{y}_q, \cdots, \boldsymbol{x}_n\otimes\boldsymbol{y}_1, \boldsymbol{x}_n\otimes\boldsymbol{y}_2, \cdots, \boldsymbol{x}_n\otimes\boldsymbol{y}_q]$$

所以

$$\text{rank}(A\otimes B)=\text{rank}(A)\text{rank}(B)=nq$$

由于 $A\otimes B$ 是 $mp\times nq$ 矩阵，因此 $A\otimes B$ 是列满秩矩阵，即 $A\otimes B$ 的列向量组 $\boldsymbol{x}_i\otimes\boldsymbol{y}_j$（$i=1, 2, \cdots, n$; $j=1, 2, \cdots, q$）是线性无关的。

反之，若列向量组 $\boldsymbol{x}_i\otimes\boldsymbol{y}_j$（$i=1, 2, \cdots, n$; $j=1, 2, \cdots, q$）是线性无关的，则 $A\otimes B$ 是列满秩的，故

$$nq=\text{rank}(A\otimes B)=\text{rank}(A)\text{rank}(B) \tag{11-4}$$

下面证明

$$\text{rank}(A)=n, \quad \text{rank}(B)=q$$

由式（11-4）知，如果 $\text{rank}(A)<n$，则必有 $\text{rank}(B)>q$，而这是不可能的，故有 $\text{rank}(A)=n$。同理，有 $\text{rank}(B)=q$。这表明矩阵 A、B 都是列满秩的，故 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 和 $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_q$ 是线性无关的。]

定理 11.1.7 设 A 为 m 阶矩阵，B 为 p 阶矩阵，则

$$|A\otimes B|=|A|^p|B|^m$$

〔证明〕 设 A 与 B 的 Jordan 标准形分别为 J_1 和 J_2，于是存在非奇异矩阵 P 与 Q，有

$$P^{-1}AP=J_1,\quad Q^{-1}BQ=J_2$$

由定理 11.1.1，有

$$A\otimes B=(P\otimes Q)(J_1\otimes J_2)(P^{-1}\otimes Q^{-1})$$

$$=(P\otimes Q)(J_1\otimes J_2)(P\otimes Q)^{-1}$$

于是 $$|A\otimes B|=|J_1\otimes J_2|$$

显然，当 J_1 和 J_2 均为下（上）三角矩阵时，$J_1\otimes J_2$ 也为下（上）三角矩阵，故有

$$|J_1\otimes J_2|=\prod_{j=1}^{p}(\lambda_1\mu_j)\prod_{j=1}^{p}(\lambda_2\mu_j)\cdots\prod_{j=1}^{p}(\lambda_m\mu_j)$$

$$=(\lambda_1\lambda_2\cdots\lambda_m)^p\left(\prod_{j=1}^{p}\mu_j\right)^m$$

$$=|A|^p|B|^m$$

其中 λ_1，λ_2，…，λ_m 为 A 的特征值，μ_1，μ_2，…，μ_p 是 B 的特征值。 ∎

定理 11.1.8 （1） 若 A、B 均为对角形矩阵，则 $A\otimes B$ 也是对角形矩阵。

（2） 若 A、B 均为对称矩阵，则 $A\otimes B$ 也是对称矩阵。

（3） 若 A、B 均为 Hermite 矩阵，则 $A\otimes B$ 也是 Hermite 矩阵。

（4） 若 A、B 均为酉矩阵，则 $A\otimes B$ 也是酉矩阵。

定理的证明留给读者。

定理 11.1.9 设 A 为 m 阶矩阵，B 为 n 阶矩阵，则存在一个 mn 阶置换矩阵（有限个初等矩阵的乘积）P，使

$$P^T(A\otimes B)P=B\otimes A$$

〔证明〕 容易验证，对矩阵 $A\otimes I_n$ 进行一系列合同变换① 可以变成 $I_n\otimes A$，即存在一个置换矩阵 P，使

$$P^T(A\otimes I_n)P=I_n\otimes A$$

又对矩阵 P，有

$$P^T(I_m\otimes B)P=B\otimes I_m$$

再由 $PP^T=I$，有

$$P^T(A\otimes B)P=P^T(A\otimes I_n)(I_m\otimes B)P$$

$$=P^T(A\otimes I_n)PP^T(I_m\otimes B)P$$

$$=(I_n\otimes A)(B\otimes I_m)$$

$$=B\otimes A$$ ∎

---

① 对矩阵的行和相应的列进行相同的初等变换叫做合同变换。例如，对调矩阵的第 i 行和第 j 行，然后再对调第 i 列和第 j 列。

最后我们指出，对 Kronecker 积也有幂的概念。记

$$A^{[k]}=\underbrace{A\otimes A\otimes\cdots\otimes A}_{k\text{个}}$$

关于 Kronecker 积的幂，我们有下面的定理。

定理 11.1.10 设 $A=A_{m\times n}$，$B=B_{n\times p}$，则

$$(AB)^{[k]}=A^{[k]}B^{[k]}$$

〔证明〕 用归纳法。当 $k=1$ 时，显然成立。设 $k-1$ 时定理成立，则

$$\begin{aligned}(AB)^{[k]}&=(AB)\otimes(AB)^{[k-1]}\\&=(AB)\otimes A^{[k-1]}B^{[k-1]}\\&=(A\otimes A^{[k-1]})(B\otimes B^{[k-1]})\\&=A^{[k]}B^{[k]}\end{aligned}$$

]

§11.2 Kronecker 积的特征值

这一节我们来讨论矩阵 A、B 的特征值与 $A\otimes B$ 的特征值间的关系。

考虑由变量 x、y 组成的复系数多项式

$$f(x,y)=\sum_{i,j=0}^{l}c_{ij}x^iy^j$$

若 A 为 m 阶矩阵，B 为 n 阶矩阵，我们来考虑由下式确定的 mn 阶矩阵

$$f(A;B)=\sum_{i,j=0}^{l}c_{ij}A^i\otimes B^j$$

例如，设 $f(x,y)=2x+xy^3$，把 $f(x,y)$ 写成

$$f(x,y)=2x^1y^0+x^1y^3$$

于是

$$f(A;B)=2A\otimes I_n+A\otimes B^3$$

特别地，若 $f(x,y)=xy$，则有

$$f(A;B)=A\otimes B$$

下面的定理给出 A、B 的特征值和 $f(A;B)$ 的特征值间的关系。

定理 11.2.1 设 λ_1，λ_2，…，λ_m 是 m 阶矩阵 A 的特征值，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，…，$\boldsymbol{x}_m$ 为 A 的对应于 $\lambda_1,\lambda_2,\cdots,\lambda_m$ 的特征向量；$\mu_1,\mu_2,\cdots,\mu_n$ 是 n 阶矩阵 B 的特征值，$\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n$ 是 B 的对应于 μ_1，μ_2，…，μ_n 的特征向量，那么 mn 个数 $f(\lambda_r,\mu_s)$（$r=1$，2，…，m，$s=1$，2，…，n）为 $f(A;B)$ 的特征值，$\boldsymbol{x}_r\otimes\boldsymbol{y}_s$ 是对应于 $f(\lambda_r,\mu_s)$ 的特征向量。

〔证明〕 由

$$A\boldsymbol{x}_r=\lambda_r\boldsymbol{x}_r,\quad B\boldsymbol{y}_s=\mu_s\boldsymbol{y}_s$$

有 $$A^i\boldsymbol{x}_r=\lambda_r^i\boldsymbol{x}_r,\quad B^j\boldsymbol{y}_s=\mu_s^j\boldsymbol{y}_s$$

于是
$$\begin{aligned}
f(A;B)\boldsymbol{x}_r\otimes\boldsymbol{y}_s &= \Big(\sum_{i,j=0}^{l} c_{ij}\ A^i\otimes B^j\Big)(\boldsymbol{x}_r\otimes\boldsymbol{y}_s)\\
&= \sum_{i,j=0}^{l} c_{ij}(A^i\otimes B^j)(\boldsymbol{x}_r\otimes\boldsymbol{y}_s)\\
&= \sum_{i,j=0}^{l} c_{ij}(A^i\boldsymbol{x}_r\otimes B^j\boldsymbol{y}_s)\\
&= \sum_{i,j=0}^{l} c_{ij}\lambda_r^i\mu_s^j\boldsymbol{x}_r\otimes\boldsymbol{y}_s\\
&= f(\lambda_r,\mu_s)\boldsymbol{x}_r\otimes\boldsymbol{y}_s
\end{aligned}$$ ▌

推论 11.2.1 $A\otimes B$ 的特征值是 mn 个数 $\lambda_r\mu_s(r=1,2,\cdots,m;s=1,2,\cdots,n)$，$\lambda_r\mu_s$ 对应的特征向量为 $\boldsymbol{x}_r\otimes\boldsymbol{y}_s(r=1,2,\cdots,m;\ s=1,2,\cdots,n)$。

推论 11.2.2 $A\otimes I_n+I_m\otimes B$ 的特征值是 $\lambda_r+\mu_s$，其对应的特征向量是 $\boldsymbol{x}_r\otimes\boldsymbol{y}_s$ $(r=1,2,\cdots,m;\ s=1,2,\cdots,n)$。

矩阵 $A\otimes I_n+I_m\otimes B$ 称为 A 和 B 的 **Kronecker 和**。

§11.3 矩阵的列展开与行展开

定义 11.3.1 设 $A=(a_{ij})_{m\times n}$，将 A 的各行依次横排得到的 mn 维行向量，称为矩阵 A 的**行展开**，记为 $\mathrm{rs}(A)$，即

$$\mathrm{rs}(A)=(a_{11},\ a_{12},\ \cdots,\ a_{1n},\ a_{21},\ a_{22},\ \cdots,\ a_{2n},\ \cdots,\ a_{m1},\ a_{m2},\ \cdots,\ a_{mn})$$

将 A 的各列依次纵排得到的 mn 维列向量，称为 A 的**列展开**，记为 $\mathrm{cs}(A)$，即

$$\mathrm{cs}(A)=(a_{11},\ a_{21},\ \cdots,\ a_{m1},\ a_{12},\ a_{22},\ \cdots,\ a_{m2},\ \cdots,\ a_{1n},\ a_{2n},\ \cdots,\ a_{mn})^T$$

由定义 11.3.1，知

$$\mathrm{rs}(A^T)=(\mathrm{cs}(A))^T$$

$$\mathrm{cs}(A^T)=(\mathrm{rs}(A))^T$$

定理 11.3.1 设 $A=(a_{ij})_{m\times n}$，$B=(b_{ij})_{n\times p}$，$C=(c_{ij})_{p\times q}$，则

$$\mathrm{rs}(ABC)=\mathrm{rs}(B)(A^T\otimes C)$$

$$\mathrm{cs}(ABC)=(C^T\otimes A)\mathrm{cs}(B)$$

〔证明〕 设 A 的第 i 行为 α_i，则 A 可以写成

$$A=\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_m\end{pmatrix}$$

于是 $$\mathrm{rs}(ABC)=(\alpha_1 BC,\ \alpha_2 BC,\ \cdots,\ \alpha_m BC)$$

设 B 的第 j 行为 β_j，C 的第 k 列为 γ_k，于是

$$\alpha_i BC=\alpha_i\begin{pmatrix}\beta_1\gamma_1 & \beta_1\gamma_2 & \cdots & \beta_1\gamma_q\\ \beta_2\gamma_1 & \beta_2\gamma_2 & \cdots & \beta_2\gamma_q\\ \cdots & \cdots & \cdots & \cdots\\ \beta_n\gamma_1 & \beta_n\gamma_2 & \cdots & \beta_n\gamma_q\end{pmatrix}$$

$$=(\beta_1,\beta_2,\cdots,\beta_n)\begin{pmatrix}a_{i1}\gamma_1 & a_{i1}\gamma_2 & \cdots & a_{i1}\gamma_q\\ a_{i2}\gamma_1 & a_{i2}\gamma_2 & \cdots & a_{i2}\gamma_q\\ \cdots & \cdots & \cdots & \cdots\\ a_{in}\gamma_1 & a_{in}\gamma_2 & \cdots & a_{in}\gamma_q\end{pmatrix}$$

$$=\mathrm{rs}(B(\alpha_i^T\otimes C))$$

故有 $$\mathrm{rs}(ABC)=\mathrm{rs}(B)(\alpha_1^T\otimes C,\ \alpha_2^T\otimes C,\ \cdots,\ \alpha_m^T\otimes C)$$
$$=\mathrm{rs}(B(A^T\otimes C))$$

同理可证 $$\mathrm{cs}(ABC)=(C^T\otimes A)\mathrm{cs}(B)$$ ▮

推论 11.3.1 设 $A=A_{m\times m}$，$B=B_{n\times n}$，$X=X_{m\times n}$，则

(1) $\mathrm{cs}(AX)=(I_n\otimes A)\mathrm{cs}(X)$

(2) $\mathrm{cs}(XB)=(B^T\otimes I_m)\mathrm{cs}(X)$

(3) $\mathrm{cs}(AX+XB)=(I_n\otimes A+B^T\otimes I_m)\mathrm{cs}(X)$

〔证明〕 (1) $\mathrm{cs}(AX)=\mathrm{cs}(AXI_n)=(I_n\otimes A)\mathrm{cs}(X)$

(2) $\mathrm{cs}(XB)=\mathrm{cs}(I_mXB)=(B^T\otimes I_m)\mathrm{cs}(X)$

(3) $\mathrm{cs}(AX+XB)=\mathrm{cs}(AX)+\mathrm{cs}(XB)$
$$=(I_n\otimes A+B^T\otimes I_m)\mathrm{cs}(X)$$ ▮

推论 11.3.2 设 $A=A_{m\times r}$，$B=B_{r\times n}$，则

$$\mathrm{rs}(AB)=\mathrm{rs}(A(I_m\otimes B))=\mathrm{rs}(B(A^T\otimes I_n))$$

〔证明〕 $$\mathrm{rs}(AB)=\mathrm{rs}(I_mAB)=\mathrm{rs}(A(I_m\otimes B))$$

$$\mathrm{rs}(AB)=\mathrm{rs}(ABI_n)=\mathrm{rs}(B(A^T\otimes I_n))$$ ▮

§11.4 线性矩阵代数方程

这一节我们讨论形如

$$A_1XB_1+A_2XB_2+\cdots+A_pXB_p=C \qquad (11\text{-}5)$$

的线性矩阵代数方程，其中 $A_j\in C^{m\times m}$，$B_j\in C^{n\times n}$（$j=1, 2, \cdots, p$），X、$C\in C^{m\times n}$

对方程（11-5）可以构造一个对应的线性方程组

$$G\boldsymbol{x}=\boldsymbol{c} \qquad (11\text{-}6)$$

其中 $\boldsymbol{x}=\mathrm{cs}(X)$，$\boldsymbol{c}=\mathrm{cs}(C)$，$G=\sum_{j=1}^{p}(B_j^T\otimes A_j)$。

定理 11.4.1 矩阵 $X\in C^{m\times n}$ 是方程（11-5）的解的充要条件是，$\boldsymbol{x}=\mathrm{cs}(X)$ 是方程（11-6）的解。

〔证明〕 对方程（11-5）施加列展开，有

$$\begin{aligned}\mathrm{cs}(C)&=\mathrm{cs}\left(\sum_{j=1}^{p}A_jXB_j\right)\\&=\sum_{j=1}^{p}\mathrm{cs}(A_jXB_j)\\&=\sum_{j=1}^{p}(B_j^T\otimes A_j)\mathrm{cs}(X)\\&=G\mathrm{cs}(X)\end{aligned}$$

即 $$G\boldsymbol{x}=\boldsymbol{c}$$

故方程（11-5）的解与方程（11-6）的相同，定理得证。 ▮

推论 11.4.1 方程（11-5）有解的充要条件是 $\mathrm{rank}(G\ \ \boldsymbol{c})=\mathrm{rank}(G)$。

推论 11.4.2 方程（11-5）有唯一解的充要条件是 G 为非奇异的。

下面我们来讨论方程（11-5）两个重要的特殊情况。

1\. 方程

$$AX+XB=C \qquad (11\text{-}7)$$

定理 11.4.2 方程（11-7）有唯一解的充要条件是 A 和 $-B$ 没有相同的特征值。

〔证明〕 与方程（11-7）对应的线性方程组是

$$\begin{aligned}\mathrm{cs}(C)&=\mathrm{cs}(AX+XB)\\&=(I_n\otimes A+B^T\otimes I_m)\mathrm{cs}(X)\end{aligned}$$

由推论 11.4.2 知，方程（11-7）有唯一解的充要条件是矩阵 $I_n\otimes A+B^T\otimes I_m$ 是非奇异的，即矩阵 $I_n\otimes A+B^T\otimes I_m$ 没有零特征值。由推论 11.2.2 知，矩阵 $I_n\otimes A+B^T\otimes I_m$ 的特征值

是 $\lambda_i+\mu_j$。于是方程（11-7）有唯一解的充要条件是 $\lambda_i+\mu_j\neq0$，即 A 与 $-B$ 没有相同的特征值。 ∎

在某种特殊情况下，我们可以得到方程（11-7）的解的表达式，即下面的定理。

定理 11.4.3 若矩阵 $A\in C^{m\times m}$、$B\in C^{n\times n}$的所有特征值具有负实部①，则方程（11-7）的唯一解 X 由下式给出

$$X=-\int_0^\infty e^{At}Ce^{Bt}dt \tag{11-8}$$

〔证明〕 由定理条件知，方程（11-7）有唯一解且 A 与 $-B$ 没有相同的特征值。 ∎

考虑由矩阵微分方程

$$\frac{dZ(t)}{dt}=AZ+ZB,\ \ Z(0)=C \tag{11-9}$$

所确定的函数矩阵 $Z(t)$。由推论 10.5.1 知，方程（11-9）的解为

$$Z(t)=e^{At}Ce^{Bt}$$

在矩阵（11-8）存在的情况下，将方程（11-9）的两端对 t 从 0 到∞积分，得

$$Z(\infty)-Z(0)=A\int_0^\infty Z(t)dt+\left(\int_0^\infty Z(t)dt\right)B$$

如果 $$Z(\infty)=\lim_{t\to\infty}e^{At}Ce^{Bt}=0 \tag{11-10}$$

那么易知式（11-8）给出的矩阵为方程（11-7）的解。

由公式（9-41）知，函数矩阵 e^{At} 可以表示成

$$e^{At}=\sum_{k=1}^{r}\sum_{j=1}^{d_k-1}t^je^{\lambda_kt}p_{kj}(A) \tag{11-11}$$

其中 λ_1，λ_2，…，λ_r 是 A 的不同的特征值，d_1，d_2，…，d_r 分别为其指数。

设 $\lambda_k=\alpha_k+i\beta_k$，则

$$e^{\lambda_kt}=e^{\alpha_kt}e^{i\beta_kt}=e^{\alpha_kt}(\cos\beta_kt+i\sin\beta_kt)$$

（$k=1, 2, \cdots, r$）。因为 $\alpha_k<0$（$k=1, 2, \cdots, r$），所以对每个 k，有

$$t\to\infty,\ \ \exp(\lambda_kt)\to0$$

从而由式（11-11），有

$$\lim_{t\to\infty}e^{At}=0$$

同理可证 $$\lim_{t\to\infty}e^{Bt}=0$$

于是，式（11-10）成立。 ∎

定理 11.4.4 设 $A\in C^{m\times m}$，$B\in C^{n\times n}$，$X\in C^{m\times n}$，则代数方程

① 如果矩阵 A 的所有特征值的实部均为负数，则称 A 是稳定的。

$$X+AXB=C \qquad (11\text{-}12)$$

有唯一解的充要条件是 $\lambda_i\mu_j \neq -1$ ($i=1, 2, \cdots, m$; $j=1, 2, \cdots, n$), λ_i、μ_j 分别为 A、B 的特征值。

〔证明〕 把方程（11-12）两端均按列展开，有

$$\begin{aligned}\mathrm{cs}(C)&=\mathrm{cs}(I_m X I_n + AXB)\\ &=(I_n\otimes I_m + B^T\otimes A)\mathrm{cs}(X)\end{aligned}$$

于是方程（11-12）有唯一解的充要条件是矩阵 $I_n\otimes I_m+B^T\otimes A$ 的特征值全不为零。由推论 11.2.1 和推论 11.2.2 知

$$1+\lambda_i\mu_j\neq 0$$

▮

（2） 方程

$$AX=XA, \quad A、X\in C^{n\times n} \qquad (11\text{-}13)$$

方程（11-13）的解 X 等价于确定所有与已知矩阵 A 可交换的矩阵。

设 $A=PJP^{-1}$，其中 J 为 A 的 Jordan 标准形，那么方程（11-13）与方程

$$JY=YJ \qquad (11\text{-}14)$$

等价，其中 $Y=P^{-1}XP$。

下面我们来考察方程（11-14）。

设 $\qquad J=\mathrm{diag}(J_1, J_2, \cdots, J_r)$

其中 $\qquad J_s=\lambda_s I_{k_s}+N_{k_s}$

I_{k_s} 为 k_s 阶单位矩阵，N_{k_s} 为对应于特征值 $\lambda_s(1\leqslant s\leqslant r)$ 的 k_s 阶 Jordan 块。

对方程（11-14）中的 Y 分块：

$$Y=\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1r}\\ Y_{21} & Y_{22} & \cdots & Y_{2r}\\ \cdots & \cdots & \cdots & \cdots\\ Y_{r1} & Y_{r2} & \cdots & Y_{rr}\end{pmatrix}, \quad Y_{st}\in C^{k_s\times k_t}$$

这样，方程（11-14）写成

$$\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r\end{pmatrix}\cdot\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1r}\\ Y_{21} & Y_{22} & \cdots & Y_{2r}\\ \cdots & \cdots & \cdots & \cdots\\ Y_{r1} & Y_{r2} & \cdots & Y_{rr}\end{pmatrix}$$

$$
=\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1r} \\ Y_{21} & Y_{22} & \cdots & Y_{2r} \\ \cdots & \cdots & \cdots & \cdots \\ Y_{r1} & Y_{r2} & \cdots & Y_{rr} \end{pmatrix}\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{pmatrix}
$$

于是有　　$J_s Y_{st} = Y_{st} J_t$，　s、$t=1, 2, \cdots, r$

或　　$(\lambda_s I_{k_s} + N_{k_s}) Y_{st} = Y_{st}(\lambda_t I_{k_t} + N_{k_t})$

即　　$(\lambda_s - \lambda_t) Y_{st} = Y_{st} N_{k_t} - N_{k_s} Y_{st}$，$(s$、$t=1, 2, \cdots, r)$

$$(11\text{-}15)$$

分两种情况

情况 1　$\lambda_s \neq \lambda_t$。用$(\lambda_s-\lambda_t)$乘方程（11-15）的两端，得

$$
\begin{aligned}
(\lambda_s-\lambda_t)^2 Y_{st} = & (Y_{st}N_{k_t} - N_{k_s}Y_{st})N_{k_t} \\
& - N_{k_s}(Y_{st}N_{k_t} - N_{k_s}Y_{st})
\end{aligned}
$$

再用（$\lambda_s-\lambda_t$）乘上式，并继续这个过程，我们有

$$
(\lambda_s-\lambda_t)^l Y_{st} = \sum_{j=0}^{l}(-1)^j\binom{l}{j}N_{k_s}^j Y_{st} N_{k_t}^{l-j} \qquad (11\text{-}16)
$$

$$(l=1, 2, \cdots,)$$

$$N_{k_s}^0 = I_{k_s}, \quad N_{k_t}^0 = I_{k_t}$$

显然，当$m \geqslant k_s$时，$N_{k_s}^m=0$。因此，在方程(11-16)中取l足够大时，有$(\lambda_s-\lambda_t)^l Y_{st}=0$。因为假设 $\lambda_s \neq \lambda_t$，故有 $Y_{st}=0$。

情况 2　$\lambda_s = \lambda_t$ 。这时方程（11-15）成为

$$
Y_{st}N_{k_s} = N_{k_t}Y_{st}, \quad 1 \leqslant s、t \leqslant r \qquad (11\text{-}17)
$$

设　　$Y_{st} = (y_{ij}) \in C^{k_s \times k_t}$

比较方程（11-17）两端矩阵的对应元素，有

$$
y_{i+1,j} = y_{i,j+1} \quad (i=1, 2, \cdots; k_s; \ j=1, 2, \cdots, k_t) \qquad (11\text{-}18)
$$

其中假设 $y_{i0} = y_{k_s+1,j} = 0$。

由式（11-18）和 k_s 与 k_t 间的大小关系，$k_s \times k_t$ 矩阵 Y_{st} 是下列三种形式之一。

（1）　若 $k_s = k_t$，则 Y_{st} 是形状如下的矩阵

$$
Y_{st} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1,k_s-1} & y_{1,k_s} \\ 0 & y_{11} & \cdots & y_{1,k_s-2} & y_{1,k_s-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & y_{11} & y_{12} \\ 0 & 0 & \cdots & 0 & y_{11} \end{pmatrix} \qquad (11\text{-}19)
$$

（2） 若 $k_s<k_t$，则 Y_{st} 是形状如下的矩阵

$$Y_{st}=(0 \quad Y_{k_s}) \tag{11-20}$$

（3） 若 $k_s>k_t$，则 Y_{st} 是形状如下的矩阵

$$Y_{st}=\begin{pmatrix} Y_{k_t} \\ 0 \end{pmatrix} \tag{11-21}$$

其中 Y_{k_s} 和 Y_{k_t} 分别为形状如（11-19）的 k_s 阶和 k_t 阶矩阵。

于是，我们得到下面的定理。

定理 11.4.5 设 $A\in C^{n\times n}$，$A=PJP^{-1}$，其中 J 为 A 的 Jordan 标准形，则矩阵 X 与 A 可交换的充要条件是 $X=PYP^{-1}$，其中 $Y=(Y_{st})$（$s=1, 2, \cdots, r$；$t=1, 2, \cdots, r$）是与 J 的 Jordan 块一致的分块矩阵。当 $\lambda_s\neq\lambda_t$ 时，$Y_{st}=0$；当 $\lambda_s=\lambda_t$ 时，Y_{st} 为式（11-19）、式（11-20）和式（11-21）三种形式之一。

例 11.4.1 设

$$A=P\left(\begin{pmatrix}2 & 1\\ 0 & 2\end{pmatrix}\dot{+}\begin{pmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{pmatrix}\dot{+}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\right)P^{-1}$$

则

$$X=P\left(\left(\begin{array}{cc:ccc} a_0 & a_1 & 0 & b_0 & b_1\\ 0 & a_0 & 0 & 0 & b_0\\ \hdashline c_0 & c_1 & d_0 & d_1 & d_2\\ 0 & c_0 & 0 & d_0 & d_1\\ 0 & 0 & 0 & 0 & d_0 \end{array}\right)\dot{+}\begin{pmatrix}e_0 & e_1\\ 0 & e_1\end{pmatrix}\right)P^{-1}$$

其中字母 a_0，a_1，b_0，b_1，c_0，c_1，d_0，d_1，d_2，e_0，e_1 表示任意复数。

显然，矩阵 $A\in C^{n\times n}$ 的任何多项式 $p(A)$ 与 A 可交换。现在问，当 A 具有什么条件时，与 A 可交换的矩阵是 A 的多项式？下面的定理给出这个问题的回答。

定理 11.4.6 每个与 $A\in C^{n\times n}$ 可交换的矩阵是 A 的多项式的充要条件是 A 的 Jordan 标准形中没有两个 Jordan 块有相同的特征值或 A 的初等因子两两互素。

〔证明〕 由定理 9.1.4 知，当 $l\leqslant n$ 时，矩阵 I，A，A^2，$\cdots$，A^{l-1} 线性无关。

如果 $AX=XA$ 意味着 X 是 A 的多项式，即 $X=p(A)$，$p(A)$ 的次数 $l\leqslant n$，且 I，A，A^2，$\cdots$，A^{l-1} 线性无关。由等式 $\alpha=\sum\limits_{s,t=1}^{r}\alpha_{st}$（见习题 11-11）知，$\alpha\geqslant n$，故必有 $\alpha=n$，即 A 的初等因子两两互素。

反之，如果 λ_1，λ_2，$\cdots$，λ_r 是 A 的不同的特征值，其次数分别为 d_1，d_2，$\cdots$，d_r。

A 的 Jordan 标准形恰好由 r 个 Jordan 块 J_i 组成，那么，方程 $AX=XA$ 的解 X 与下述形式矩阵的直和相似

$$
Y_i=\begin{pmatrix} y_0^{(i)} & y_1^{(i)} & \cdots & y_{d_i-1}^{(i)} \\ & y_0^{(i)} & \cdots & y_{d_i-2}^{(i)} \\ & & \ddots & \vdots \\ & & & y_0^{(i)} \end{pmatrix}
$$

$(i=1, 2, \cdots, r)$。由方程 (9-12) 可知，多项式 $p(\lambda)$ 满足条件

$$
p(\lambda_i)=y_0^{(i)},\ p'(\lambda_i)=y_1^{(i)},\ \cdots,\ \frac{1}{(d_i-1)!}p^{(d_i-1)}(\lambda_i)=y_{d_i-1}^{(i)},
$$

$(i=1, 2, \cdots, r)$，且有

$$
\begin{aligned}
X&=P\mathrm{diag}(Y_1, \cdots, Y_r)P^{-1}\\
&=P\mathrm{diag}(p(J_1), \cdots, p(J_r))P^{-1}\\
&=Pp(J)P^{-1}=P(PJP^{-1})
\end{aligned}
$$

其中 $p(\lambda)$ 的次数不超过 $n-1$。 ▮

把定理 11.4.3 应用到矩阵

$$
\begin{pmatrix} I_m & X \\ 0 & I_n \end{pmatrix} \text{和} \begin{pmatrix} A & 0 \\ 0 & -B \end{pmatrix}
$$

可以得到下面的定理。

定理 11.4.7 设 $A\in C^{m\times m}$，$B\in C^{n\times n}$，$A=PJ_1P^{-1}$，$B=QJ_2Q^{-1}$，其中

$$
J_1=\mathrm{diag}(\lambda_1 I_{k_1}+N_{k_1}, \cdots, \lambda_p I_{k_p}+N_{k_p})
$$
$$
J_2=\mathrm{diag}(\mu_1 I_{t_1}+N_{t_1}, \cdots, \mu_q I_{t_q}+N_{t_q})
$$

分别是 A、B 的 Jordan 标准形。于是方程

$$
AX+XB=0
$$

的任何解具有形式

$$
X=PYP^{-1}
$$

其中

$$
Y=\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1q} \\ Y_{21} & Y_{22} & \cdots & Y_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ Y_{p1} & Y_{p2} & \cdots & Y_{pq} \end{pmatrix}
$$

是方程 $\qquad J_1Y+YJ_2=0$

的解，$Y_{ij}\in C^{k_i\times k_j}$（$s=1, 2, \cdots, p$；$t=1, 2, \cdots, q$）。因此 Y 具有这样的性质：当 $\lambda_s\neq\mu_t$ 时，$Y_{st}=0$；当 $\lambda_s=-\mu_t$ 时，Y_{st} 是式（11-19）、（11-20）、（11-21）三种形式之一。

习　　题

11-1　证明对任意矩阵 $A=(a_{ij})_{n\times n}$，有

（1）　　$I_n\otimes A=\mathrm{diag}(A, A, \cdots, A)$

（2）　　$$A\otimes I_m=\begin{pmatrix} a_{11}I_m & a_{12}I_m & \cdots & a_{1n}I_m \\ \cdots & \cdots\cdots & \cdots & \cdots \\ a_{n1}I_m & a_{n2}I_m & \cdots & a_{nn}a_m \end{pmatrix}$$

（3）　　$I_m\otimes I_n=I_{mn}$

11-2　证明下列等式：

（1）　　$k(A\otimes B)=kA\otimes B=A\otimes kB$，　k 为任意复数

（2）　　$(A+B)\otimes C=A\otimes C+B\otimes C$

（3）　　$(A\otimes B)\otimes C=A\otimes(B\otimes C)$

11-3　证明推论 11.1.1。

11-4　证明定理 11.1.2。

11-5　证明定理 11.1.8。

11-6　证明推论 11.2.1。

11-7　证明推论 11.2.2。

11-8　设 $A\in C^{m\times m}$，证明对于在 A 的谱上定义的函数 $f(\lambda)$，有

（1）　　$f(I_n\otimes A)=I_n\otimes f(A)$

（2）　　$f(A\otimes I_n)=f(A)\otimes I_n$

11-9　设 $C\in C^{mn\times mn}$ 是 $A\in C^{m\times n}$ 和 $B\in C^{n\times n}$ 的 Kronecker 和，证明 $e^C=e^A\otimes e^B$。

11-10　证明：如果 A 和 B 是正定的，则 $A\otimes B$ 也是正定的。

11-11　设 α_{st} 为初等因子（$\lambda-\lambda_s$）^{k_s} 和（$\lambda-\lambda_t$）^{k_t}（$1\leqslant s$、$t\leqslant r$）的最大公因式的次数。证明：$AX=XA$，A、$X\in C^{n\times n}$ 中的 X 有

$$a=\sum_{t,s=1}^{r}\alpha_{st}$$

个不定元素。

（提示：考察如果 $\lambda_s=\lambda_t$，则 $\alpha_{st}=\min(k_s, k_t)$）。

11-12　设$A \in C^{m \times m}$，$B \in C^{n \times n}$，$C \in C^{m \times n}$。如果 A 和 B 的谱半径分别为 μ_A 和 μ_B，且 $\mu_A \mu_B < 1$，则方程

$$X = AXB + C$$

有唯一解，且 X 由下式给出：

$$X = \sum_{i=0}^{\infty} A^i C B^i$$

第十二章 广义逆矩阵

本章讨论应用广泛的广义逆矩阵的基本内容。所谓广义逆矩阵是普遍意义上的逆矩阵的推广，它是由求解线性方程的问题而引进的。

本章讨论的内容有：广义逆矩阵的概念及基本性质，自反广义逆矩阵，伪逆矩阵，A^+ 的各种表示，广义逆矩阵在线性方程组和矩阵方程中的应用。

§12.1 广义逆矩阵的概念及其性质

我们知道，如果A 为非奇异的，则线性方程组

$$A\boldsymbol{x}=\boldsymbol{b}$$

有唯一解 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$ 。在一般情况下，上述方程组有解的充要条件是

$$\mathrm{rank}(A\boldsymbol{b})=\mathrm{rank}(A)$$

但这时是否也能用某个矩阵 G 把解表示成 $\boldsymbol{x}=G\boldsymbol{b}$ 的形式呢？下面我们来讨论这个问题。

设线性方程组

$$A\boldsymbol{x}=\boldsymbol{b} \tag{12-1}$$

其中 $A\in C^{m\times n}$，$\boldsymbol{b}\in C^{m}$，$\boldsymbol{x}\in C^{n}$。

定义 12.1.1 对于 $\mathrm{rank}(A\boldsymbol{b})=\mathrm{rank}(A)$ 的任意 $\boldsymbol{b}$，设 $\boldsymbol{x}$ 是方程（12-1）的解，当有使$\boldsymbol{x}=A^-\boldsymbol{b}$ 成立的矩阵 A^- 存在时，称 A^- 为 A 的**广义逆矩阵**。

在 $\mathrm{rank}(A\boldsymbol{b})=\mathrm{rank}(A)$ 的条件下，方程（12-1）有解，但其解不是唯一的，所以一般说来 A^- 也不是唯一的。

下面我们来讨论广义逆矩阵的一些性质。

命题 12.1.1 A^- 存在的充要条件是有满足 $AA^-A=A$ 的 $A^-\in {}^{n\times m}$ 存在。

〔证明〕 如果 A^- 存在，则在方程（12-1）有解时，$\boldsymbol{b}$ 可以表示成$\boldsymbol{b}=A\boldsymbol{z}$，，$\boldsymbol{z}\in C^{n}$，而且只能这样表示。设 $A\boldsymbol{x}=\boldsymbol{b}$ 的解可以表示成 $A^-\boldsymbol{b}$，则 $AA^-\boldsymbol{b}=\boldsymbol{b}$，即 $AA^-A\boldsymbol{z}=A\boldsymbol{z}$。因该式对任意的 $\boldsymbol{z}\in C^{n}$ 均成立，故$AA^-A=A$。

反之，设 $\boldsymbol{x}\in C^{n}$为$A\boldsymbol{x}=\boldsymbol{b}$ 的解，则 $AA^-A\boldsymbol{x}=A\boldsymbol{x}$，即 $AA^-\boldsymbol{b}=\boldsymbol{b}$，于是 $\boldsymbol{x}=A^-\boldsymbol{b}$ 也是解。]

命题 12.1.2 对任意的 $A\in C^{m\times n}$，至少有一个 A^- 存在，而且总有 $\mathrm{rank}(A^-)\geqslant \mathrm{rank}(A)$。

〔证明〕 若 $A=0$，则对任意的 $X\in C^{n\times m}$ 满足

$$AXA=A$$

若 $A\neq 0$，$\mathrm{rank}(A)=r(\geqslant 1)$。将 A 表示成 $A=BC$，$B\in C^{m\times r}$，$C\in C^{r\times n}$，$\mathrm{rank}(B)=\mathrm{rank}(C)$。

令 $$A^{+}=C^{H}(CC^{H})^{-1}(B^{H}B)^{-1}B^{H}$$

则有

$$AA^{+}A=A$$

于是 A^{+} 是 A 的一种广义逆矩阵。

而且，因$\mathrm{rank}(A)=\mathrm{rank}(AA^{-}A)$，则$\mathrm{rank}(A)=\mathrm{rank}(A^{-})$ ∎

由命题 12.1.1 和命题 12.1.2 知，对于任意矩阵 A，有定义 12.1.1 意义上的广义逆矩阵A^{-}存在。作为和定义 12.1.1 等价的定义，所谓 A 的广义逆矩阵也可以定义成使得

$$A=AA^{-}A \tag{12-2}$$

成立的矩阵 A^{-}，于是当 A 为非奇异矩阵时，有 $A^{-}=A^{-1}$，且仅限于 A^{-1}。

根据上述等价定义，易知

$$(A^{T})^{-}=(A^{-})^{T},\ (A^{H})^{-}=(A^{-})^{H} \tag{12-3}$$

$$(AA^{-})^{2}=AA^{-},\ (A^{-}A)^{2}=A^{-}A \tag{12-4}$$

命题 12.1.3 $AA^{-}B=B \Leftrightarrow \exists C$，使$B=AC \Leftrightarrow \exists D$，使 $B=AA^{H}D$。

〔证明〕 如果 $AA^{-}B=B$，令 $C=A^{-}B$ 即知有第一个等价性；反之，$AA^{-}B=AA^{-}AC=AC=B$。

第二个等价性，令$D=(A^{+})^{H}C$即可得证，其中 A^{+} 是在命题 12.1.2 的证明中所用过的 A 的广义逆矩阵。如后所述，对于 A 是唯一确定的，而且具有：$(A^{+}A)^{H}=A^{+}A$。反之，令 $A^{H}D=C$ 即可。 ∎

命题 12.1.4 对任意矩阵 C 和任意正定 Hermite 矩阵 B，有

$$(AA^{H}+B)(AA^{H}+B)^{-}AC=AC \tag{12-5}$$

〔证明〕 由 $B\geqslant 0$，则 B 可以写成

$$B=DD^{H}$$

于是

$$AA^{H}+B=[A\quad D]\begin{bmatrix}A^{H}\\ D^{H}\end{bmatrix}$$

由命题 12.1.3，以

$$AC=[A\ \ D]\begin{bmatrix}C\\ 0\end{bmatrix}$$

可以得到有使下式成立的矩阵 M 存在

$$AC=[A\ \ D]\begin{bmatrix}A^{H}\\ D^{H}\end{bmatrix}M=(AA^{H}+B)M$$

因此 $$(AA^{H}+B)(AA^{H}+B)^{-}AC$$

$$=(AA^H+B)(AA^H+B)^-(AA^H+B)M$$

$$=(AA^H+B)M=AC$$]

作为（12-5）式的特殊情况，若令 $B=0$，$C=I_n$，则有

$$(AA^H)(AA^H)^-A=A \qquad (12-6)$$

同时，无论 AA^H 的广义逆矩阵定成什么样，$A^H(AA^H)^-A$ 都不变，而且是Hermite矩阵（见习题 12-2）。因此，取式（12-6）的转置，有

$$A^H(AA^H)^-(AA^H)=A^H \qquad (12-7)$$

命题 12.1.5 设 $U\in C^{n\times m}$ 为任意矩阵，A^- 是 A 的一个广义逆矩阵，则

$$X=A^-+U-A^-AUAA^- \qquad (12-8)$$

是 A 的广义逆矩阵，而且 A 的任何广义逆矩阵都可以表示成这种形式。

〔证明〕 在式（12-8）两端分别同时左乘和右乘 A，因

$$AXA=AA^-A+AUA-AA^-AUAA^-A$$

$$=A+AUA-AUA=A$$

由命题 12.1.1 知，X 是 A 的广义逆矩阵。当 X 为 A 的任意广义逆矩阵时，由

$$A(X-A^-)A=0$$

可知
$$X=A^-+(X-A^-)-A^-A(X-A^-)AA^-$$

成立。令 $U=X-A^-$，则式（12-8）成立。]

命题 12.1.6 设 A^- 为 A 的一个广义逆矩阵，则对任意的 V、$W\in C^{n\times m}$，有

$$X=A^-+V(I_m-AA^-)+(I_n-A^-A)W \qquad (12-9)$$

是 A 的广义逆矩阵，而且 A 的任意广义逆矩阵都可以用该形式表示。

〔证明〕 由式（12-9）易知

$$AXA=A$$

而且当 X 是 A 的任意广义逆矩阵时，令

$$V=X-A^-,\ W=XAA^-$$

则对于 V、W，式（12-9）成立。]

§12.2 自反广义逆矩阵

从上节的讨论中可以看到，广义逆矩阵 A^- 和普通逆矩阵 A^{-1} 有若干共同的性质，同时也有许多不同之处。例如，A^{-1} 只有在 A 是方阵且 A 为满秩时才存在，而 A^- 总是存在的。

对于 A^{-1}，有 $(A^{-1})^{-1}=A$，与此相对应，在 $AA^-A=A$，$A^-AA=A^-$ 成立的条件下，考虑 A 的广义逆矩阵，为此，我们有下面的定义。

定义12.2.1 对于$A\in C^{m\times n}$，使

$$AA^{-}A=A,\quad A^{-}AA^{-}=A^{-}\tag{12-10}$$

成立的矩阵$A^{-}\in C^{n\times m}$称为A的**自反广义逆矩阵**，记作A_r^{-}，即

$$AA_r^{-}A=A,\quad A_r^{-}AA_r^{-}=A_r^{-}\tag{12-11}$$

命题12.2.1 设A^{-}为A的某个广义逆矩阵，则A^{-}是自反的充要条件是$\mathrm{rank}(A^{-})=\mathrm{rank}(A)$。

〔证明〕 如果A^{-}是自反的，由式（12.10），有

$$\mathrm{rank}(A^{-})\leqslant\mathrm{rank}(A)$$

$$\mathrm{rank}(A)\leqslant\mathrm{rank}(A^{-})$$

故 $$\mathrm{rank}(A^{-})=\mathrm{rank}(A)$$

反之，设 $\mathrm{rank}(A^{-})=\mathrm{rank}(A)$，由式（12-10），有

$$\mathrm{rank}(AA^{-})=\mathrm{rank}(A^{-})$$

于是，对某个$\boldsymbol{x}\in C^{m}$，$(AA^{-})\boldsymbol{x}=0$意味着$A^{-}\boldsymbol{x}=0$，即

$$(AA^{-}(I_m-AA^{-}))=0$$

从而

$$A^{-}(I_m-AA^{-})=0$$ ∎

命题12.2.2 设$A\neq0$，当$\mathrm{rank}(A)=r$时，将A分解成$A=BC$，其中$B\in C^{m\times r}$，$C\in C^{r\times n}$，$\mathrm{rank}(B)=\mathrm{rank}(C)=r$，则

$$A_r^{-}=C_R^{-1}B_L^{-1}\tag{12-12}$$

是A的自反广义逆矩阵，而且任何自反广义逆矩阵都可以用该形式表示。其中B_L^{-1}，C_R^{-1}分别为B和C的左逆矩阵($B_L^{-1}B=I_r$)和右逆矩阵($CC_R^{-1}=I_r$)。

〔证明〕 由式（12-12），有

$$AA_r^{-}A=BCC_R^{-1}B_L^{-1}BC=BC=A$$

同样有 $$A_r^{-}AA_r^{-}=A_r^{-}$$

于是式（12-12）的A_r^{-}为A的自反广义逆矩阵。

设A_r^{-}为A的任意自反广义逆矩阵，则

$$AA_r^{-}A=A$$

即 $$BCA_r^{-}BC=BC$$

将上式两端分别左乘B_L^{-1}和右乘C_R^{-1}，有

$$CA_r^{-}B=I_r$$

上式意味着$A_r^{-}B$是C的右逆矩阵，CA_r^{-}是B的左逆矩阵。因此可以写成

$$A_I^- B = C_R^{-1},\quad CA_I^- = B_L^{-1}$$

将上式代入 $A_I^- = A_I^- A A_I^-$ 中，得

$$A_I^- = A_I^- BCA_I^- = C_R^{-1} B_L^{-1}$$ ∎

这里我们指出，因为 B、C 有最大秩，则 B_L^{-1}、C_R^{-1} 必定存在。例如 $(B^H B)^{-1} B^H$、$C^H(CC^H)^{-1}$ 分别为 B_L^{-1} 和 C_R^{-1} 中的一个。因为 B_L^{-1}、C_R^{-1} 存在，则对任意 A、A_I^- 也存在（但不唯一）。

§12.3 伪逆矩阵

当 A 用最大秩的积 $A=BC$ 表示时，可以证明

$$A^+ = C^H(CC^H)^{-1}(B^H B)^{-1} B^H \tag{12-13}$$

是 A 的自反广义逆矩阵。需要指出的是，尽管 B、C 不是由 A 唯一确定，但是由式(12-13)给出的 A^+ 却是唯一的①。矩阵 A^+ 称为 A 的伪逆矩阵或 Moore-Penrose 逆矩阵。

由 A^+ 的定义，易知

$$(A^+)^+ = A \tag{12-14}$$

$$(A^H)^+ = (A^+)^H,\quad (AA^H)^+ = (A^+)^H A^+ \tag{12-15}$$

命题 12.3.1

$$(AA^+)^H = AA^+,\quad (A^+A)^H = A^+A \tag{12-16}$$

〔证明〕 由式(12-13)，有

$$AA^+ = BCC^H(CC^H)^{-1}(B^H B)^{-1} B^H$$

$$= B(B^H B)^{-1} B^H$$

$$A^+A = C^H(CC^H)^{-1}(B^H B)^{-1} B^H BC$$

$$= C^H(CC^H)^{-1} C$$

于是 $$(AA^+)^H = AA^+,\quad (A^+A)^H = A^+A$$ ∎

A^+ 除了是自反广义逆矩阵外，还满足命题 12.3.1，即 A^+ 由下面四个条件唯一确定

$$\begin{cases} AA^+A = A,\quad A^+AA^+ = A^+ \\ (AA^+)^H = AA^+,\quad (A^+A)^H = A^+A \end{cases} \tag{12-17}$$

式(12-17)也可作为 A^+ 的定义。

命题 12.3.2

$$A^+ = (A^H A)^+ A^H = A^H (AA^H)^+ \tag{12-18}$$

---

① 对于 $A = 0_{m\times n}$ 定义 $A^+ = 0_{n\times m}$

〔证明〕 设 $A=BC$，则 A^HA 的最大秩分解可写成

$$A^HA=C^H(B^HBC)$$

于是

$$(A^HA)^+=(B^HBC)^H(B^HBCC^HB^HB)^{-1}(CC^H)^{-1}C$$
$$=C^H(CC^H)^{-1}(B^HB)^{-1}(CC^H)^{-1}C$$

因此

$$(A^HA)^+A^H=C^H(CC^H)^{-1}(B^HB)^{-1}(CC^H)^{-1}C(C^HB^H)$$
$$=A^+$$

同样可证

$$A^H(AA^H)^+=A^+$$

命题 12.3.3 设 $A\in C^{m\times n}$，$B\in C^{n\times p}$，令 $B_1=A^+AB$，$A_1=AB_1B_1^+$，则

$$(AB)^+=B_1^+A_1^+ \qquad (12\text{-}19)$$

〔证明〕 因为

$$A_1B_1=AB_1B_1^+B_1=AB_1=AA^+AB=AB$$

于是只要证明 $B_1^+A_1^+$ 是 A_1B_1 的伪逆矩阵即可。

设 $X=B_1^+A_1^+$，$Y=A_1B_1$。利用关系 $A_1B_1=AB_1=AB$，容易证明

$$XYX=X,\quad YXY=Y$$

若能证明 YX、XY 是 Hermite 矩阵，由式(12-17)则可证明 $X=Y^+$。

在

$$YX=A_1(B_1B_1^+A_1^+)=AB_1B_1^+(B_1B_1^+A_1^+)$$
$$=AB_1B_1^+A_1^+=A_1A_1^+$$

中，因 $A_1A_1^+$ 是 Hermite 矩阵，故 YX 也是 Hermite 矩阵。

$$XY=B_1^+(A_1^+A_1)B_1=B_1^H(A_1^+A_1)^HB_1$$
$$=B_1^+(A_1^+A_1B_1B_1^+)^HB_1$$
$$=B_1^+(B_1B_1^+A_1^+A_1)B_1$$
$$=B_1^+(A^+ABB_1^+A_1^+A_1)B_1$$

同样

$$XY=B^+B_1B_1^+B_1$$

这就证明了 XY 是 Hermite 矩阵。 ∎

§12.4 A⁺的各种表示

对给定的矩阵 $A\in C^{m\times n}$，由式(12-13)可以求出 A^+。下面我们再给出 A^+ 的几种表示法。

（1） 当 A 为对角矩阵，即$A=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$时，易知

$$A^{+}=\mathrm{diag}(\beta_1, \beta_2, \cdots, \beta_n)$$

其中$\beta_i=\lambda_i^{-1}(\lambda_i \neq 0)$，$\beta_i=0$ $(\lambda_i=0)$。

（2） 我们把 Hermite 矩阵 $A^H A$ 表示成

$$A^H A=UMU \tag{12-20}$$

其中 U 是酉矩阵，它的列由 $A^H A$ 的特征向量构成。而且 M 是以 $A^H A$ 的特征值为主对角线上元素的实对称矩阵。因为 $A^H A$ 是准正定的，所以它的特征值均为非负。设 M^+ 如上述那样确定，则

$$(A^H A)^{+}=UM^{H}U \tag{12-21}$$

成立。此外，将式（12-21）代入式（12-18）便可求出 A^+。

如果

$$M=\begin{pmatrix} \lambda_1 & & & & & & \\ & \lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix} \tag{12-22}$$

则式（12-20）和式（12-21）可以写成谱分解形式。

$$A^H A=\sum_{j=1}^{r}\lambda_j^2 E_j \tag{12-23}$$

$$(A^H A)^{+}=\sum_{j=1}^{r}\lambda_j^{-2} E_j \tag{12-24}$$

其中$E_j^2=E_j(j=1, 2, \cdots, r)$，$E_j E_k=0(j \neq k)$。现引入多项式

$$p_i(\lambda)=(\lambda-\lambda_1)^2\cdots(\lambda-\lambda_{i-1})^2(\lambda-\lambda_{i+1})^2\cdots(\lambda-\lambda_r)^2$$

则

$$p_i(A^H A)=\sum_{j=1}^{r}p_i(\lambda_j^2)E_j \tag{12-25}$$

因$p_i(\lambda_j^2)=0$（对任意 $j \neq i$），故式（12-25）变成

$$p_i(A^H A)=p_i(\lambda_i^2)E_i$$

由该式解出 E_i 后代入式（12-23）、（12-24），便可得到 A^+ 的表示式：

$$A^{+}=\sum_{j=1}^{r}\lambda_j^{-2}E_j A^H$$

$$= \sum_{i=1}^{r} \lambda_i^{-2} \frac{p_i(A^H A)}{p_i(\lambda_i^2)}$$

$$= \sum_{i=1}^{r} \lambda_i^{-2} \frac{\prod_{i \neq j} (A^H A - \lambda_j^2 I)}{\prod_{i \neq j} (\lambda_i^2 - \lambda_j^2)} A^H \tag{12-26}$$

（3） 利用 A 的所谓奇异值分解，也可以得到 A^+ 的表示式.

在式（12-20）和式（12-22）中，令

$$U = [P \quad \bar{P}] \tag{12-27}$$

$$L = \begin{bmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_r^2 \end{bmatrix} \tag{12-28}$$

其中P为 $n\times r$ 矩阵，$\bar{P}$ 为 $n\times(n-r)$ 矩阵，则式（12-20）可以写成

$$A^H A = UMU^H = PLP^H \tag{12-29}$$

因为 U 是酉矩阵，所以 $P^H P = I_r$。设 L 的正定平方根矩阵为 $L^{1/2}$，则显然 $PL^{1/2}P^H$ 是 $A^H A$ 的准正定平方根矩阵，并且易知，和式（12-21）一样，其伪逆矩阵为

$$[(A^H A)^{1/2}]^+ = PL^{1/2}P^H \tag{12-30}$$

此外，由习题12-4的（3）、（4）、（5）有

$$\begin{aligned} A^+ A &= (A^H A)^+ (A^H A) \\ &= [(A^H A)^{1/2}]^+ (A^H A)^{1/2} \end{aligned} \tag{12-31}$$

将式（12-30）和式（12-31）代入 $A = AA^+A$，有

$$\begin{aligned} A &= A[(A^H A)^{1/2}]^+ (A^H A)^{1/2} \\ &= A[(A^H A)^{1/2}]^+ PL^{1/2}P^H \\ &= Q^H L^{1/2} P^H \end{aligned} \tag{12-32}$$

其中 $$Q^H = A[(A^H A)^{1/2}]^+ P \tag{12-33}$$

式（12-32）称为 A 的奇异值分解形式。

由式（12-33），得

$$\begin{aligned} Q^H LQ &= A[(A^H A)^{1/2}]^+ PLP^H [(A^H A)^{1/2}]^+ A^H \\ &= A[(A^H A)^{1/2}]^+ A^H A [(A^H A)^{1/2}]^+ A^H \\ &= A(A^+ AA^+ A)A^H = AA^H \end{aligned} \tag{12-34}$$

由式（12-30）和式（12-34）知，关于 Q 有下列性质：

$$QQ^H = P^H((A^HA)^{1/2})^+ A^H A((A^HA)^{1/2})^+ P$$

$$= P^H(PL^{-1/2}P^H)A^H A(PL^{-1/2}P^H)P$$

$$= I_r \qquad (12\text{-}35)$$

再者，因奇异值分解也是A的最大秩分解，由式（12-32）可求出A^+，即设$C = L^{1/2}P^H$，$B = Q^H$，则由式（12-13）A^+可表示成

$$A^+ = (PL^{1/2})(L^{1/2}P^HPL^{1/2})^{-1}(QQ^H)^{-1}Q$$

$$= PL^{-1/2}Q \qquad (12\text{-}36)$$

其中$P \in C^{n\times r}$，$Q \in C^{r\times m}$是满足下列关系的矩阵

$$A^HA = PLP^H,\quad P^HP = I_r$$

$$AA^H = Q^HLQ,\quad QQ^H = I_r$$

综上所述，我们有下面的定理。

定理12.4.1 当$\text{rank}A = r$时，设A^HA的特征值（同样，或AA^H的特征值）为λ_1^2，λ_2^2，…，λ_r^2，这时，A^+可用式（12-26）表示；当$L = \text{diag}(\lambda_1^2, \lambda_2^2, \cdots, \lambda_r^2) > 0$，$L^{1/2} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r) > 0$时，$A^+$可用式（12-36）表示。

（4） 最后，我们再给出A^+的一种常用的表示式。先来证明一个引理。

引理12.4.1 对于任意Hermite矩阵$A \in C^{n\times n}$，有

$$AA^+(=A^+A) = \lim_{\delta\to 0}(A+\delta I)^{-1}A$$

$$= \lim_{\delta\to 0} A(A+\delta I)^{-1}$$

〔证明〕 对于足够小的$|\delta|>0$，因$A+\delta I$是非奇异的，故$(A+\delta I)^{-1}$存在。又A是Hermite矩阵，则可由酉矩阵U表示成

$$A = UMU^H$$

$$M = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

其中λ_1，λ_2，…，λ_n为A的特征值。

由 $A = AA^+A = A^2A^+$，知

$$(A+\delta I)^{-1}A = (A+\delta I)^{-1}A^2A^+$$

$$= U(M+\delta I)^{-1}M^2U^HA^+$$

再由 $$\lim_{\delta\to 0}(M+\delta I)^{-1}M^2$$

$$= \lim_{\delta\to 0}\text{diag}((\lambda_1+\delta)^{-1}\lambda_1^2, \cdots, (\lambda_n+\delta)^{-1}\lambda_n^2)$$

$$= \text{diag}(\lambda_1, \cdots, \lambda_n)$$

$$= M$$

有 $$\lim_{\delta\to 0}(A+\delta I)^{-1}A=UMU^HA^+=AA^+$$

同样可以求出

$$\lim_{\delta\to 0}A(A+\delta I)^{-1}=AA^+$$ ∎

定理 12.4.2

$$A^+=\lim_{\delta\to 0}(A^HA+\delta^2I_n)^{-1}A^H$$

$$=\lim_{\delta\to 0}A^H(AA^H+\delta^2I_m)^{-1} \tag{12-37}$$

（证明） 由

$$A^H=A^H(A^H)^+A^H=A^H(AA^+)^H=A^HAA^+$$

有 $$\lim_{\delta\to 0}(A^HA+\delta^2I_n)^{-1}A^H$$

$$=\lim_{\delta\to 0}(A^HA+\delta^2I_n)^{-1}A^HAA^+$$

由引理 12.4.1 和习题 12-4 中的(5)，有

$$\lim_{\delta\to 0}(A^HA+\delta^2I_n)^{-1}A^HAA^+$$

$$=(A^HA)(A^HA)^+A^+=A^+AA^+=A^+$$

于是 $$\lim_{\delta\to 0}(A^HA+\delta^2I_n)^{-1}A^H=A^+$$

同理可证

$$\lim_{\delta\to 0}A^H(AA^H+\delta^2I_m)^{-1}=A^+$$ ∎

§12.5 在线性方程组中的应用

这一节我们来讨论用 A^- 表示方程组 $Ax=b$ 的一般解和在 $\mathrm{rank}(Ab)\neq\mathrm{rank}(A)$ 的情况下，用 A^+ 给出在某种意义上最好的近似解。

定理 12.5.1 设线性方程组

$$Ax=b,\ A\in C^{m\times n},\ b\in C^m \tag{12-38}$$

则

$Ax=b$ 有解时

（1） $Ax=b$ 有解当且仅当 $AA^-b=b$。

（2） 设 A^- 为 A 的任意广义逆矩阵，将它固定时，解的一般形式可以表示成

$$x=A^-b+(I_n-A^-A)z \tag{12-39}$$

其中 $z\in C^n$ 是任意向量。

（3） $\widetilde{x}=A^{+}b$ 是 $Ax=b$ 的所有解中有最小范数（对于欧氏范数）的解。

$Ax=b$ 无解时

（4） $\widetilde{\widetilde{x}}=A^{+}b$ 是所有最小二乘近似解中有最小范数的解，即在满足下列关系

$$\|Ax-b\|\leqslant\|Az-b\|,\ 对\ \forall z\in C^{n} \tag{12-40}$$

的最小二乘近似解 x 的集合中，对于 $\widetilde{\widetilde{x}}=A^{+}b$，

$$\|\widetilde{\widetilde{x}}\|\leqslant\|x\| \tag{12-41}$$

成立，其中$\|\cdot\|$为欧氏范数。

〔证明〕（1） 当 $Ax=b$ 对某个 $x\in C^{n}$ 成立时，有

$$AA^{-}b=AA^{-}Ax=Ax=b$$

反之，由 $\quad AA^{-}b=b$

有 $x=A^{-}b$ 满足 $Ax=b$。

（2） 在式(12-39)两端左乘以 A，有

$$Ax=AA^{-}b+A(I_{n}-A^{-}A)z=AA^{-}b$$

由（1），因 $AA^{-}b=b$，显然式（12-39）的 x 是式(12-38)的解，而且，当 $\widetilde{x}$ 是任意解时，若令 $z=\widetilde{x}-A^{-}b$，则

$$\begin{aligned}(I_{n}-A^{-}A)z&=(I_{n}-A^{-}A)(\widetilde{x}-A^{-}b)\\&=\widetilde{x}-A^{-}b-A^{-}A\widetilde{x}+A^{-}AA^{-}b\\&=\widetilde{x}-A^{-}b-A^{-}b+A^{-}b\end{aligned}$$

于是式（12-39）成立。

（3）、（4） 用欧氏范数时

$$\begin{aligned}\|Az-b\|^{2}&=\|(AA^{+}-I_{m})b+A(z-A^{+}b)\|^{2}\\&=\|(AA^{+}-I_{m})b\|^{2}+\|A(z-A^{+}b)\|^{2}\\&\quad+b^{H}(AA^{+}-I_{m})^{H}A(z-A^{+}b)\\&\quad+(z-A^{+}b)^{H}A^{H}(AA^{+}-I_{m})b\end{aligned}$$

由式（12-16）有 $(AA^{H})A=A$，$A^{H}(AA^{+})=A$，将其代入上式消去最后两项，得

$$\begin{aligned}\|Az-b\|^{2}&=\|(AA^{+}-I_{m})b\|^{2}+\|A(z-A^{+}b)\|^{2}\\&\leqslant\|(AA^{+}-I_{m})b\|^{2},\ 对\ \forall z\in C^{n},\ b\in C^{m}\end{aligned}$$

特别当 b 固定时，若选取 z 使得 $Az=AA^{+}b$，则 $\|Az-b\|^{2}$ 与其最小值 $\|(AA^{+}-I_{m})b\|^{2}$ 一致。此外，由（2），设 $w\in C^{n}$ 为任意向量，则 $Az=AA^{+}b$ 的一般解可以表示成

$$\begin{aligned}z&=A^{+}(AA^{+}b)+(I_{n}-A^{+}A)w\\&=A^{+}b+(I_{n}-A^{+}A)w\end{aligned}$$

但由于 $$\|z\|^2=\|A^+\boldsymbol{b}\|^2+\|(I_n-A^+A)w\|^2 \qquad (12\text{-}42)$$

因此，在使 $\|Az-\boldsymbol{b}\|^2$ 为最小的 z 中，$\tilde{z}=A^+\boldsymbol{b}$ 有最小范数。当然，在 $Ax=\boldsymbol{b}$ 有解的情况下，由(1)，因为 $AA^+\boldsymbol{b}=\boldsymbol{b}$ 成立，所以 $\tilde{z}=A^+\boldsymbol{b}$ 是满足 $Az=\boldsymbol{b}$ 的 z 中有最小范数的向量。

§12.6　在矩阵方程AXB=C中的应用

这一节我们来考察矩阵方程 $AXB=C$ 的解，其中 $A\in C^{m\times n}$，$B\in C^{p\times q}$，$C\in C^{m\times q}$，$X\in C^{n\times p}$。

矩阵方程 $AXB=C$ 是线性方程 $Ax=\boldsymbol{b}$ 的扩展形。当 $p=q=1$，$B=1$ 时，即为方程 $Ax=\boldsymbol{b}$。

定理 12.6.1　(1)　$AXB=C$ 有解的充要条件是 $AA^-CB^-B=C$。

(2)　$AXB=C$ 的一般解可由下式给出：

$$X=A^-CB^-+(Z+A^-AZBB^-)，\text{对}\forall Z\in C^{n\times p}$$

(3)　在 $AXB=C$ 的解中，$\tilde{X}=A^+CB^+$ 有最小范数，其中矩阵范数是 $\|X\|^2=\mathrm{tr}(X^HX)$。

(4)　在 $AXB=C$ 无解的情况下，在所有最小二乘近似解中，$\tilde{X}=A^+CB^+$ 具有最小范数，即令

$$M=\{X\mid \|AXB-C\|\leqslant\|AZB-C\|，\text{对}\forall Z\in C^{n\times p}\}$$

则当 $\tilde{X}_1\in M$，且对 $\forall X\in M$ 时，均有

$$\|\tilde{X}\|\leqslant\|X\|$$

其中假定 $\|X\|^2=\mathrm{tr}(X^HX)$。

〔证明〕(1)、(2)的证明留给读者。现在证明(3)、(4)。

先考虑

$$A=(a_{11}\ \ a_{12}),\quad B=\begin{pmatrix} b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}$$

$$C=(c_{11}\ \ c_{12}),\quad X=\begin{pmatrix} x_{11} & x_{12}\\ x_{21} & x_{22}\end{pmatrix}$$

的情况。这时方程 $AXB=C$ 可以写成

$$\begin{pmatrix} a_{11}b_{11} & a_{12}b_{11} & a_{11}b_{21} & a_{12}b_{21}\\ a_{11}b_{12} & a_{12}b_{12} & a_{11}b_{22} & a_{12}b_{22}\end{pmatrix}\begin{pmatrix} x_{11}\\ x_{21}\\ x_{12}\\ x_{22}\end{pmatrix}=\begin{pmatrix} c_{11}\\ c_{12}\end{pmatrix}$$

记X和C的列展开为$\boldsymbol{x}$和$\boldsymbol{c}$，即

$$\boldsymbol{x}=[x_{11}\ x_{21}\ x_{12}\ x_{22}]^T,\quad \boldsymbol{c}=[c_{11}\ c_{12}]^T$$

上面关于X的元素的线性方程组可以写成

$$(B^T\otimes A)\boldsymbol{x}=\boldsymbol{c} \tag{12-43}$$

在一般情况下，也完全一样，将 X 和 C 的元素按上述原则排成列向量，则 $AXB=C$ 可以写成式（12-43）的形式。由定理 12.4.1，有

$$\tilde{\boldsymbol{x}}=(B^T\otimes A)^+\boldsymbol{c} \tag{12-44}$$

是（12-43）的最小二乘近似解中有最小范数的解，即对于满足

$$\|(B^T\otimes A)\boldsymbol{x}-\boldsymbol{c}\|^2\leqslant\|(B^T\otimes A)\boldsymbol{z}-\boldsymbol{c}\|^2,\ 对\forall \boldsymbol{z}\in C^{n\times p} \tag{12-45}$$

的向量 $\boldsymbol{x}$

$$\|\tilde{\boldsymbol{x}}\|^2\leqslant\|\boldsymbol{x}\| \tag{12-46}$$

成立。这里所用的向量范数是欧氏范数，则

$$\begin{aligned}\|\boldsymbol{x}\|^2&=(x_{11}^2+x_{12}^2+\cdots+x_{1n}^2+\cdots+x_{n1}^2+x_{n2}^2+\cdots+x_{nn}^2)\\&=\mathrm{tr}(X^HX)\end{aligned} \tag{12-47}$$

成立。因此，若式（12-45）、式（12-46）按着 $\boldsymbol{x}\to X$，$\boldsymbol{c}\to C$ 改写，则分别变成

$$\|AXB-C\|^2\leqslant\|AZB-C\|^2,\ 对\ \forall Z\in C^{n\times p} \tag{12-48}$$

$$\|\tilde{X}\|^2\leqslant\|X\|^2 \tag{12-49}$$

式中$\|X\|^2=\mathrm{tr}(X^HX)$。最后，因$(B^T\otimes A)^+=(B^T)^+\otimes A^+$，故 $\tilde{\boldsymbol{x}}=(B^T\otimes A)^+\boldsymbol{c}$ 变成

$$\tilde{X}=A^+CB^+$$

在 $AXB=C$ 有解的情况下，由式(12-48)，对于满足 $AXB=C$ 的X，$\tilde{X}$ 是满足式(12-49)的解。 ∎

习　　题

12-1　证明式（12-3）和式（12-4）。

12-2　证明 $0_{m\times n}$ 的自反广义逆矩阵仅为 $0_{n\times m}$。

12-3　证明式（12-14）和式（12-15）。

12-4　证明下列等式

（1）　$(A^HA)^+=A^+(A^H)^+$，$(AA^H)^+=(A^H)^+A^+$

（2）　$(A^HA)^+=A^+(AA^H)^+A=A^H(AA^H)^+(A^H)^+$

（3）　$AA^+=(AA^H)(AA^H)^+=(AA^H)^+(AA^H)$

（4） $A^+A=(A^HA)(A^HA)^+=(A^HA)^+(A^HA)$

（5） 若 $A^H=A$，则 $(A^2)^+=(A^+)^2$，$A^2(A^2)^+=(A^2)^+A^2=AA^+$。

（6） 若 $A^H=A$，则 $AA^+=A^+A$。

12-5 证明$(A\otimes B)^+=A^+\otimes B^+$。

12-6 当U是Hermite矩阵，M是对角矩阵时，试用式(12-28)证明$(UMU^H)^+=UM^+U^H$成立。

12-7 令$U=V(I-AA^-)+(I-A^-A)W$，试证式(12-9)可以写成式(12-8)的形式。

12-8 证明定理12.6.1的(1)、(2)。

名 词 索 引

八 划

九 划

十 划

十 一 划

十 二 划

十　三　划

十　四　划

参 考 文 献

1. 须田信英等，自动控制中的矩阵理论.
2. P.Lancaster and M.Tismenetsky,The Theory of Matrices Second Edition with Applications（1985）.
3. R.Bellman,Introduction to Matrix Analysis（1970）.
4. A.Berman and J.Plemmons,Nonnegative Matrices In The Mathematical Sciences（1979）.
5. I.Gohberg, P.Lancaster and Rodman Matrix Polynomials（1982）.